



University of Tennessee, Knoxville
**TRACE: Tennessee Research and Creative
Exchange**

[Doctoral Dissertations](#)

[Graduate School](#)

12-2015

The Roles of Self-Affirmation and Introspection in Correction for Automatic Prejudice

Kevin Lee Zabel

University of Tennessee - Knoxville, kzabel1@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss



Part of the [Social Psychology Commons](#)

Recommended Citation

Zabel, Kevin Lee, "The Roles of Self-Affirmation and Introspection in Correction for Automatic Prejudice. " PhD diss., University of Tennessee, 2015.
https://trace.tennessee.edu/utk_graddiss/3624

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Kevin Lee Zabel entitled "The Roles of Self-Affirmation and Introspection in Correction for Automatic Prejudice." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Experimental Psychology.

Michael A. Olson, Major Professor

We have read this dissertation and recommend its acceptance:

Lowell A. Gaertner, Garriy Shteynberg, Victor E. Ray

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

The Roles of Self-Affirmation and Introspection in Correction for Automatic Prejudice

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Kevin Lee Zabel

December 2015

“You can accomplish anything by working together.”

- Grandpa Chuck Zabel Sr.

I dedicate this work to my wonderful wife, Barbara, who has provided unwavering love and support over the last five years.

I dedicate this work to my loving family, including my father (Chuck), mother (Tammy), brother (Keith), and sister (Nicole), as well as my grandmothers (Ruth Ann and Donna) and grandfather (Chuck), who have always supported me in all my endeavors. Time together may be harder to come by with time, but it makes us appreciate the time we do get to share together all that much more. I dedicate this to Kim Peck and Bruce Eisenmenger. Thank you for always believing in me and supporting me, as well as everything you have done for me over the last several years. I truly cannot thank you enough.

I dedicate this work to my mentors, including Michael Olson, Drew Christopher, and Scott Chesebro, who were instrumental in setting the course of my current professional goals and aspirations, as well as providing unrelenting support to me, both professionally and personally.

I dedicate this work to my extended family, work colleagues, and friends, including Lane Chesebro, Joe Salvatore, Amy Heger, Dawn Howerton, Rick Kendrick, Joy Phillips, and the dozens of undergraduate research assistants and students I have had the pleasure of working with at the University of Tennessee. Without your support, this work would not have been possible, nor as enjoyable of an endeavor.

I dedicate this work to *all* that work hard, love often, and do unto others as they would have done unto them in all walks of life, that together make this world a better place.

Acknowledgements

Here, I acknowledge the love and support of many loved ones who have made this work possible.

Barbara, thank you for providing unwavering support, rain or shine, over the last five years. Being alongside you in life makes everything better and makes me feel more like “myself,” and I cannot wait for the adventures we are able to share together as we enjoy life together, hand in hand.

I thank my loving family, including my father (Chuck Zabel Jr.), mother (Tammy Zabel), brother (Keith Zabel), and sister (Nicole Zabel), as well as my grandmother (Ruth Ann Zabel) and grandfather (Chuck Zabel Sr.). I thank my father for always putting family first and for sacrificing very much in order for me to have opportunities he never had. You are an inspirational father and person to me, and I love you very much. Thank you to my mother for providing love and care throughout my life. Thank you to my sister, Nicole, for always being supportive, caring, and loving toward myself, family, and friends. Nicole, you are a selfless, compassionate person, and I am grateful to have you in my life. Thank you to my brother and identical twin, Keith, for being by my side every part of life. Keith, we have done everything together since we were children, and now we even professionally are part of the same field. I would not have it any other way. Thank you for always being solid in your support toward me, for being honest and candid with me, and for being your jovial, witty self. I love both you and Lauren very much. Thank you to my Grandma Zabel and Grandpa Zabel for providing steadfast support for me throughout my life. Whether it be caring for me on Saturdays, preparing amazing meals, playing euchre while watching sports, or literally building a barn to play basketball in, thank you sincerely for everything you have done for me. Your house is always a refuge during stressful times, and I appreciate your sage advice, as well as unwavering support so much. Coffee always tastes best at your kitchen table with you.

I am forever grateful for the guidance provided by my mentors, advisors, and friends, including Drew Christopher, Scott Chesebro, and Michael Olson. You each have been instrumental in setting the course of my current professional goals and aspirations, as well as in providing unrelenting support to me, both professionally and personally.

Drew, if it were not for you, I never would have had the guidance necessary to be prepared for graduate school. As an undergraduate, I had no idea what kind of opportunities were available in graduate school, nor that it even existed as an option. You literally opened up a new world for me professionally. I am grateful for the opportunities you provided me to get involved in research within your lab, as well as to present research at conferences and in paper form as publications. I thank you for the trust you afforded me and the potential that you saw in me, even when I did not. I sincerely thank you for your mentorship, and I am grateful that our friendship exists to this day. In addition to Drew, I thank the entire Department of Psychological Science at Albion

College for their fantastic undergraduate training and preparation, as well as the opportunity to work alongside them as a Visiting Professor of Psychology during the Spring 2014 semester. Specifically, thank you to Mareike Wieth, Jacque Carlson, Holger Elischberger, Andrea Francis, Jeff Wilson, Barbara Keyes, Eric Hill, Tammy Jechura, Ed Bujdos, and Schara Swan for their support and trust in my teaching capabilities. It was truly a joy to have the opportunity to teach back at my alma mater, and it rekindled the love I had for Albion College and the department of psychological science faculty members and staff. I am truly grateful for the opportunity.

Scott, thank you for opening my mind to different perspectives and cultures, and literally “opening my eyes” to seeing the same world through a different perceptual lens. The Chicago Center experience was empowering in every way, including living in an apartment with others, navigating an international city on my own, working for a major United States financial company, and approaching and embracing diversity for the first time in my life. My experiences at the Chicago Center not only shaped my current research interests in racial prejudice, but provided me the courage to pursue my graduate studies and professional livelihood in regions of the country that were previously unknown to me. These experiences have of course deepened my understanding of the world and appreciation for different perspectives, as well as led to a more enriching life, all of which would not have been possible without the Center’s influence. I feel fortunate that our friendship continues to this day. Thank you furthermore for allowing me the opportunity to work with you as an apprentice. Not only did this deepen my resolve to pursue a psychology degree, but it also provided me invaluable professional experience that has been useful in my work as a graduate research assistant and instructor, as well as in starting a tenure-track assistant professorship at Western New England University. You have helped me work through and finalize important decisions in my life (e.g., deciding on the University of Tennessee for graduate school), and it was only fitting that you married Barbara and me earlier this fall. Thank you very much for everything- I am forever grateful.

Michael, thank you so much for everything over the last five years. There were times throughout when self-doubt would seep in, but your unwavering support and encouragement sustained me throughout. Thank you for challenging me, and for providing steadfast support in my pursuit of my teaching and research goals. You are an amazingly selfless individual, and that you were willing to support my teaching goals, which sometimes curtailed my research productivity, is evidence of your selflessness and concern for student well-being. Thank you for literally always having your door open to discuss matters, large and small, or to chat about research-related questions. You have influenced not only my research prowess, but also the support I strive to provide my students in teaching and research. Finally, your ability to strive toward work-life balance, and to always put your family first while conducting research and teaching at such a high level has always personally been an inspiration to me. I am grateful for the opportunity to work with you and to have you as my mentor and friend. Thank you sincerely for everything, Michael. I cannot thank you enough.

I also want to thank my extended family, work colleagues, and friends, including Lane Chesebro, Joe Salvatore, Matt Shaffer, Amy Heger, Dawn Howerton, Rick Kendrick, Joy Phillips, Tiffany Smith, and Jordan Troisi. Without your support, this work would not have been possible. Thank you so much for all your support, whether it be emotionally, in aiding in research directly, or in listening to my concerns and worries when I needed a listening ear. I appreciate everything so very much.

Thank you to the Department of Psychology at the University of Tennessee. Specifically, a heartfelt thank you goes out to Lowell Gaertner, Garriy Shteynberg, Jeff Larsen, Erin Hardin, Connie Ogle, and Mary Ellen Hunsberger. Thank you all for your support in research, teaching, and professional development. I am extremely fortunate to have learned from all of you. Thank you additionally to Victor Ray and Joe Miles, who provided invaluable service on my dissertation and masters' thesis committees, respectively. I sincerely appreciate your work and willingness to serve on the committee. Furthermore, thank you to the family of Newton W. and Wilma C. Thomas for their generous support in the form of a research fellowship during the 2010-2014 academic years. I am forever grateful for your financial support.

A heartfelt and sincere thank you also goes out to all the students I was fortunate enough to teach at the University of Tennessee in Knoxville, as well as work with side by side on research. Specifically, thank you to Garren Hoskins, Sean Purdy, Quinton Thompson, Kayleigh Hickerson, Alysha Moncus, Margaret Turner, Maddie Bainer, Kaitlin Hudson, Leah Ortiz, Julia Collier, Emily Bazel, Sarah Solovey, Kevin Woods, Danielle Salton, Lisa Nagel, Kristin Samples, Meredith Farnsworth, Lauren Lynch, Anthony Gardner, Carly Oliver, Caitlin Brandow, Tyler Maybrier, Janie Brackett, Lauren Duren, Alex Bucceli, and Melissa Freeman for their invaluable service and aid in conducting empirical research projects in the Olson Social Cognition Lab alongside me, as well as other graduate students. I am especially thankful for Carly Oliver's aid in coordinating every step of the data collection process for my graduate research projects during the Spring 2014 semester. Carly, your work was invaluable, and performed in an extremely competent manner- I hope you know how much I genuinely appreciate all the tremendous work you did.

The bottom line is that I am forever grateful for everyone's support. The support of the individuals above has provided me the opportunity to pursue aspirations and opportunities I would have otherwise never dreamed possible. I am fortunate to have had and currently have all of you in my life, and I will always appreciate your love and support.

Abstract

Egalitarian-oriented Whites tend to employ the strategy of “liking everyone,” as opposed to correcting for their automatic prejudices, as a means of avoiding prejudiced reactions (Zabel & Olson, 2014). Congruent with motivational theoretical perspectives regarding prejudice (i.e., Aversive Racism; Gaertner & Dovidio, 1986), I contend that a lack of introspection into one’s automatic prejudices due to a self-image threat may be driving this tendency. In the experiment I report here, I assessed the automatic racial attitudes of egalitarian- (high Concern) and conflict avoidance-motivated (high Restraint) Whites. Then, participants were randomly assigned to introspect (or not) on their automatic racial biases, as well as to self-affirm (or not) using a values importance task. Following, participants completed an impression formation task in which they rated Black and White targets. Results indicated that egalitarian-oriented (high Concern) individuals with pro-White automatically-activated attitudes corrected for their prejudices by providing greater pro-Black trait ratings when self-affirmation preceded racial attitude introspection. This pattern of prejudice correction did not occur among high Concern individuals who introspected on their racial attitudes prior to being self-affirmed, or did not introspect or self-affirm at all. Findings are congruent with previous research attesting to the utility of self-affirmation in increasing attention to otherwise self-threatening information, and provide insights as to how the strategies employed by aversive racists and their negative implications may be reduced.

Table of Contents

Chapter 1: Introduction	1
Dual Process Theories	2
The MODE Model	4
Restraint and Concern Motivation	5
Target Construal, Motivation to Control Prejudiced Reactions, and Prejudice ... Correction	8
Empirical Examinations of Target Construal Effects in Bias Correction	10
Theoretical Perspectives Underlying a Lack of Bias Correction	13
Introspection	14
Introspection Versus Awareness	15
Lacking the Skill to Introspect	17
Motivational Influences on Introspection	19
Self-Affirmation Theory.....	25
Chapter 2: Hypotheses	28
Chapter 3: Method	30
Participants	30
Materials and Procedure	30
Motivation to control prejudiced reactions (MCPR)	30
Private self-consciousness	31
Public self-consciousness	31
Automatic prejudice	31
Self-affirmation manipulation	32

Introspection manipulation	33
Impression formation task	34
Pro-black/anti-black attitudes	35
Subtle and blatant prejudice	36
Symbolic racism	36
Self-affirmation manipulation coding	37
Introspection manipulation coding	37
Chapter 4: Results	38
Data Preparation	38
Motivation to control prejudiced reactions	38
Automatic prejudice	38
Trait ratings	39
Self-affirmation manipulation coding	41
Introspection manipulation coding	41
Preliminary Analyses	42
Primary Analyses	44
Analysis strategy	44
Admission of racial biases	44
Prejudice correction on trait ratings	46
Prejudice correction regarding explicit attitude judgments	52
Role of restraint motivation in trait judgment prejudice correction	55
Role of restraint motivation in self-reported racial attitude prejudice correction	59

Supplementary Analyses	60
Trait rating response latencies	60
Chapter 5: Discussion	66
References	81
Appendices	97
Appendix A: Tables.....	98
Appendix B: Figures	105
Vita	112

List of Tables

Table 1: Descriptive statistics and correlations between variables for all participants (averaging across condition)	99
Table 2: Descriptive statistics and correlations between variables for control condition participants.....	100
Table 3: Descriptive statistics and correlations between variables for introspection only condition participants.....	101
Table 4: Descriptive statistics and correlations between variables for self- affirmation only condition participants.....	102
Table 5: Descriptive statistics and correlations between variables for introspection first/self-affirmation second condition participants	103
Table 6: Descriptive statistics and correlations between variables for self- affirmation first/introspection second condition participants	104

List of Figures

Figure 1: Concern motivation x automatically-activated racial attitudes interaction in predicting overall pro-Black trait ratings in the self-affirmation first/introspection second condition	106
Figure 2: Concern motivation x automatically-activated racial attitudes interaction in predicting overall pro-Black trait ratings in the control condition	107
Figure 3: Concern motivation x automatically-activated racial attitudes interaction in predicting overall pro-Black trait ratings in the introspection first/self-affirmation second condition	108
Figure 4: Concern motivation x automatically-activated racial attitudes interaction in predicting mean overall critical target trait rating response latencies in the self-affirmation first/introspection second condition	109
Figure 5: Concern motivation x automatically-activated racial attitudes interaction in predicting overall mean critical target trait rating response latencies in the control condition	110
Figure 6: Concern motivation x automatically-activated racial attitudes interaction in predicting overall mean critical target trait rating response latencies in the self-affirmation only condition	111

Chapter 1: Introduction

Racial prejudices and stereotypes are pervasive in American society and learned via imitations and observations of parents, teachers, and peers from an early age. Furthermore, they can influence behavior and judgments swiftly and automatically (Brewer, 1988; Fiske & Neuberg, 1990; Greenberg & Pyszczynski, 1985). A person does not have to actively endorse a prejudice or stereotype in order for it to have a direct effect on judgments or behavior (Correll, Park, Judd, & Wittenbrink, 2002). Institutional rules and national laws, as well as the development of social norms that punish and reprimand obvious expressions of discrimination have shifted the nature of prejudices from more overt antipathy to more covert and subtle forms. Not surprisingly, given this shift, measures of prejudice and theoretical perspectives regarding their composition and expression have evolved as well.

Indeed, egalitarian-oriented motives and motives to avoid the social costs of appearing prejudiced have been found to moderate the influence of automatic prejudice on judgments and behaviors (Dunton & Fazio, 1997; Olson & Fazio, 2004; see Fazio & Olson, 2014 for a review). Egalitarian-oriented individuals still possess automatic racial prejudices but may not wish to look inward and examine their biases, due to a threat to their egalitarian threat image. This is consistent with Aversive Racism Theory (Gaertner & Dovidio, 1986), which contends egalitarian-minded, automatically prejudiced individuals engage in specific strategies to avoid situations, contexts, attributions, and processes that may reveal their automatic prejudices. Consistent with Aversive Racism Theory, recent research indicates that egalitarian-oriented (but automatically prejudiced)

individuals tend to follow a ‘like everyone’ heuristic instead of attempting to correct for the influence of their prejudices on judgments and behaviors (Zabel & Olson, 2014).

Thus, ironically, the very individuals who would seemingly desire most to avoid prejudiced reactions may be unlikely to introspect and subsequently correct for their automatic racial biases due perhaps to the potential threat to their egalitarian self-image. The present research investigates whether a lack of introspection may be facilitating bias correction failure among egalitarian-oriented Whites, as well as how these individuals might be prompted to introspect on their prejudices and potentially correct for the effects of their automatic biases in their judgments and behaviors.

Dual-Process Theories

Dual-process theories explain mental processes that underlie social behavior as stemming from an interplay between automatic and controlled processes. Automatic mental processes are deemed to influence behavior in a more unconscious, effortless, and involuntary manner, relative to more controlled processes, which influence behavior in a more conscious, effortful, and voluntary manner. Dual process theories have enjoyed abundant empirical support and extension in several social psychology realms, including social perception, attitudes, prejudice, cognition, and the self (see Sherman, Gawronski, & Trope, 2014 for a review). For instance, in the social perception realm, models of impression formation (Brewer, 1988; Fiske & Neuberg, 1990) distinguish between impressions formed via a more initial automatic, unconscious, and involuntary category judgment, relative to a more controlled, conscious, voluntary, and effortful individualized judgment. Spontaneous trait inferences by which traits are inferred directly based on

behavior consist of both automatic and controlled processes that can be activated simultaneously (McCarthy & Skowronski, 2014). The Associative-Propositional Evaluation (APE) Model distinguishes evaluative responses toward attitude objects that are based more in associative, relative to propositional processes (Gawronski & Bodenhausen, 2006). Whether evaluations activated by associative processes are *consistent* with propositional reasoning in a given context influences the degree to which more controlled processes are activated to influence judgments and behavior.

The Elaboration Likelihood Model (Petty & Cacioppo, 1986) is one of the more classical dual process theories, and distinguishes between persuasive attitude change that occurs via a less effortful, less elaborative, less motivated (peripheral) route relative to a more effortful, more elaborative, and more motivated (central) route. Motivation, as well as the availability and use of cognitive resources are requirements for increased message elaboration fundamental to persuasive attitude change via the central route to persuasion.

Not only can automatic mental components such as stereotypes and automatic prejudice influence explicit attitudes, judgments, and behaviors (Fazio & Olson, 2014; Sherman et al., 2014), but ambivalence caused by a mismatch between more explicitly generated attitudes and automatic mental processes can lead to greater deliberation and scrutiny (under conditions of high cognitive resources; Bodenhausen, Mccrae, & Sherman, 1999). Greater deliberation, scrutiny, and controlled processing is important to combatting the influence of automatic mental processes (i.e., stereotypes) on judgments and behaviors (Clark, Wegener, Brinol, & Petty, 2009), correcting for prejudices, and ultimately, reducing the expression of prejudices. In the present work, I consider what

may prompt egalitarian-oriented but automatically prejudiced individuals to notice their prejudices and then correct for their influence in judgments and behaviors.

The MODE Model

The MODE model emphasizes the interplay between more controlled and automatic processes in attitude-behavior relations, as well as specifies the conditions under which bias correction is likely to occur (Fazio, 1990). Although the Mode Model could be applied to any context to understand the interplay of attitudes and motivation in predicting relevant judgments and behaviors (e.g., romantic relationships; McNulty, Olson, Meltzer, & Shaffer, 2013), it has been primarily applied to the dynamics of racial prejudice (see Fazio & Olson, 2014 for a review). The MODE Model aims to determine whether attitudes predict judgments and behavior in a more spontaneous (i.e., automatic), relative to more deliberative (controlled) manner. The MODE Model argues that automatic attitudes toward an attitude object are automatically-activated upon perception of that object. In the absence of motivation (e.g., accuracy, accountability, concern with social desirability, to control prejudice) and the opportunity (time and cognitive resources) for more motivated, controlled processes to operate, automatic attitudes are predicted to directly inform judgments and behaviors toward an attitude object. Applied to prejudice, under conditions of depleted cognitive resources or low motivation to control prejudice, automatically prejudiced individuals should favor White (relative to Black) targets behaviorally, as well as in terms of judgments and impressions.

However, when the opportunity to devote mental resources exists, motivated processes are able to exert an influence on judgments and behaviors that may steer

judgments and behavior away from those implied by one's automatically-activated racial attitude. For instance, when sufficient time and cognitive resources exist, individuals highly motivated to control their prejudiced reactions can correct for their automatic racial attitudes. In instances of high motivation to avoid prejudice, individuals will rely on their own naïve subjective theories to determine their bias levels and to calibrate their behaviors and judgments (Wegener & Petty, 1997; Wegener, Petty, & Dunn, 1998; Wilson & Brekke, 1994). Of course, individuals may under-, appropriately, or over-correct for their automatic biases in a given context, which could yield a variety of positive or negative consequences. However, when sufficient time and cognitive resources exist, as well as motivation to control prejudiced reactions, the tendency is that as prejudiced automatic racial attitudes increase, so too do *pro-Black* judgments and behaviors (Bodenhausen, Todd, & Richeson, 2009; Chien, Wegener, Petty, & Hsiao, 2014; Dunton & Fazio, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Olson & Fazio, 2004; Sanbonmatsu & Fazio, 1990; Schuette & Fazio, 1995; Towles-Schwen & Fazio, 2003; Zabel & Olson, 2014). Thus, highly motivated White individuals correct for their automatic racial prejudices by favoring Black (relative to White) targets.

Restraint and Concern Motivation

The effects of two factors of motivation to control prejudiced reactions have been well-documented in bias correction from a MODE Model perspective using Dunton and Fazio's (1997) motivation to control prejudices (MCPR) questionnaire. The first factor, Concern With Appearing Prejudiced (henceforth referred to as *Concern*) consists of a personal desire to appear unprejudiced to oneself and to others, and includes items like "I

feel guilty when I have a negative thought or feeling about a Black person.” Concern is positively correlated ($r = .50$) with humanitarianism-egalitarianism (Fazio & Hilden, 2001) and internal motivation ($r = .38$) to control prejudice (IMS; Plant & Devine, 1998). IMS consists of a strong desire to avoid prejudiced reactions driven by internal forces and egalitarian values, as opposed to the potential social costs and norm-violation conflict linked with appearing prejudiced (Plant & Devine, 1998). IMS is linked with active efforts to facilitate positive interracial interactions (e.g., high interest in engaging with partner; Plant, 2004), as well as the goal pursuit of approaching egalitarianism (Plant, Devine, & Peruche, 2010). Indeed, in contrast to actively engaging in strategies to *avoid* the expression of prejudice (i.e., avoiding stereotypes, avoiding talking about sensitive topics, avoiding behavior that could be construed as prejudiced, keeping the conversation short), high IMS individuals tend to focus more on *approaching* interracial interactions with behaviors (i.e., smiling, asking questions, making eye contact, sharing self-information) congruent with their egalitarian self-image. Concern’s correlation with IMS suggests that it too facilitates more of an approach-focus to interracial interactions.

High Concern is associated with reports of more positive intergroup contact with Black individuals in childhood (Towles-Schwen & Fazio, 2001). Moreover, high Concern individuals feel guilty when they violate internal standards regarding the expression of prejudices (Fazio & Hilden, 2001) and seem motivated to redress past racial inequalities (Fazio & Olson, 2014). In sum, research indicates that high Concern individuals have a strong egalitarian focus and desire to redress past inequalities which promotes an approach focus to interracial interaction. Such individuals, to the extent that

they harbor automatic racial biases, are akin to aversive racists (Gaertner & Dovidio, 1986; Song Hing, Chung-Yan, Hamilton, & Zanna, 2008).

The second factor of the MCPR questionnaire, Restraint to Avoid Dispute (henceforth referred to as *Restraint*) entails a desire to avoid race-related conflict with or about Black individuals, and includes items like, “If I were participating in a class discussion and a Black student expressed an opinion with which I disagreed, I would be hesitant to express my own viewpoint.” Restraint is linked with external motivation ($r = .35$) to control prejudice (EMS; Plant & Devine, 1998), which is a strong desire to avoid prejudiced reactions due to the social costs associated with violating social norms that discourage the expression of prejudices. EMS is linked with efforts to avoid prejudices in interracial interactions (e.g., avoiding interracial interactions; Plant, 2004), as well as the goal pursuit of avoiding prejudice (Plant et al., 2010). Instead of *approaching* interracial interactions with egalitarian-motivated behaviors, high EMS individuals actively engage in strategies to *avoid* the expression of prejudice (i.e., avoiding talking about sensitive topics, avoiding behavior that could be construed as prejudiced) to avoid the conflict and social costs associated with violating egalitarian social norms. Restraint’s correlation with EMS (and not IMS) suggests that it too facilitates more of an avoidance-focus to interracial interactions (Fazio & Olson, 2014).

High Restraint is associated with reduced and less positive contact with Blacks in childhood (Towles-Schwen & Fazio, 2001). Unlike Concern, Restraint is not tied to egalitarianism (Fazio & Hilden, 2001), indicating that for high Restraint individuals, the desire to avoid prejudice is more tied to potential social costs or ramifications than

violating an egalitarian-minded self-image. Both Concern and Restraint, when activated, moderate the impact of automatic prejudice on race-related judgments and behaviors. However, as I explain next, these motives may each be especially applicable, activated, and applied to prejudice correction in specific contexts depending on how social targets are construed.

Target Construal, Motivation to Control Prejudiced Reactions, and Prejudice

Correction

Intriguing trends have been observed with regard to the influence of Concern and Restraint in leading to correction for the effect of automatically-activated prejudice in behaviors and judgments. One point has to do with the contexts in which Restraint and Concern typically lead to corrective effects. Specifically, target construal may play an important role in whether Concern or Restraint is activated and in turn, leads to corrections for automatic prejudice. Social targets can be flexibly construed at the individual level, the category level, or somewhere in between (Brewer, 1988; Fiske, Lin, & Neuberg, 1999; Hamilton & Sherman, 1996; Macrae & Bodenhausen, 2000). I contend that it is likely that some perceiver motives are more relevant (and hence more influential) when social targets are construed at the individual level, whereas other motives are more relevant to impression formation when targets are construed at the category level. It is reasoned that Restraint, in being more conflict-avoidance-oriented, is more likely to be activated when social targets are construed at the individual level, since conflict is more likely to arise in interpersonal, individual-level interactions with others. Concern, in being more egalitarian-oriented and focused on equal treatment toward

Blacks, is more likely to be salient, activated, and operative in group contexts (Fazio & Olson, 2014). Thus, target construal may influence whether Restraint or Concern is activated in a given context, and in turn, moderates the effects of automatic prejudice on judgments and behaviors.

In support, Restraint has often yielded automatic prejudice correction effects when social targets are construed at the individual, relative to category level. In one study (Olson & Fazio, 2004), and consistent with the MODE Model (Fazio & Olson, 2014), automatically-activated attitudes directly predicted trait ratings toward Black and White targets among low Restraint individuals. However, among those high in Restraint, a pattern of correction emerged in which prejudiced Whites were more likely to give positive trait ratings of Black (relative to White) targets. A similar pattern of prejudice correction emerged among high Restraint individuals when asked to “describe their impressions of the typical Black male undergraduate” (Dunton & Fazio, 1997, pp. 322). In both cases, Restraint moderated the automatic attitudes-impressions relationship in contexts in which social targets were being construed at the individual level.

In comparison, Concern has demonstrated automatic prejudice correction effects when social targets are construed at the category (group), relative to individual level. For instance, Concern moderated the relationship between automatic prejudice and responses on the Modern Racism Scale (MRS), a measure of prejudice toward Blacks as a group (Fazio et al., 1995, see also Payne, Cheng, Govorun, & Stewart, 2005). Automatic prejudice directly predicted MRS scores for those low in Concern, but higher Concern participants corrected for their prejudice in responding to the MRS (Dunton & Fazio,

1997). As previously stated, we have reasoned that because Restraint is more conflict-avoidance oriented, and Concern more egalitarian-oriented, that each is more likely to be activated in specific target construal contexts congruent with their underlying tenets. The findings highlighted above suggested that construal level might activate specific goal-related motives that in turn would influence the expression of automatically-activated attitudes on behaviors and impressions.

Empirical Examinations of Target Construal Effects in Bias Correction

Recently, I experimentally examined these possibilities (Zabel & Olson, 2014). Specifically, participants for whom Concern and Restraint scores were available were administered Fazio et al.'s (1995) affective priming task to assess automatic prejudice toward Blacks. Then, participants completed a task designed to encourage construal of social targets either at the individual or the category level. Following, participants provided impressions of Black and White targets as in previous research (Olson & Fazio, 2004).

As expected, and consistent with the MODE Model (Fazio & Olson, 2014), automatic prejudice directly influenced impressions of Black relative to White targets among those less motivated to avoid prejudice. Correction for automatic prejudice was observed in the impressions of more motivated participants, but the application of different motives to control prejudice (i.e., Restraint and Concern) was dependent on target construal level. As expected, Restraint (and not Concern) led to more corrected impressions among participants construing targets at the individual level. That is, as expected, as Restraint increased, participants in the individual target construal condition

provided increasingly positive impressions toward Black (relative to White) targets as their automatic prejudice levels increased. The moderational role of Restraint in the automatic attitudes-trait impressions relationship both replicates and extends previous research (Dunton & Fazio, 1997; Olson & Fazio, 2004) by demonstrating that Restraint is especially likely to be activated and operational in automatic prejudice correction within individual (but not category) target construal contexts.

Unexpectedly, a pattern of automatic prejudice correction among those high in Concern failed to emerge among those in the category construal condition. However, Concern was positively associated with trait ratings of *both* Black and White targets, as well as a relative preference for Black over White targets, albeit not significantly. As Concern increased, individuals tended to adopt a “like everyone” strategy when forming impressions of social targets and rated both Black and White targets as high in desirable traits. These correlations between Concern and trait ratings toward Black and White targets are in line with Concern’s focus on positive group treatment and an egalitarian orientation (Fazio & Hilden, 2001). The perplexing finding was that Concern increases did not lead to proportionately greater pro-Black trait ratings as automatically-prejudiced attitudes increased, as had been demonstrated by Restraint when social targets were construed at the individual level. In other words, high Concern individuals failed to correct for their automatic prejudice (Zabel & Olson, 2014).

These findings were replicated and extended in a second study (Zabel & Olson, 2014) using a forced-choice impression formation paradigm in which participants directly compared Black and White targets. In this experiment, participants had to indicate a

preference between two simultaneously presented White and Black targets with regard to each of five desirable traits. I reasoned that once the “like everyone” option was revoked, and participants were forced to favor one social target over the other, high Concern individuals would be more likely to correct for their automatic prejudices in the category construal condition. Indeed, I expected that correcting for prejudices by favoring Black (relative to White) social targets would be more justifiable when participants were forced to favor one target over another.

However, results were consistent with previous findings (Zabel & Olson, 2014): Restraint moderated the relationship between automatic attitudes and trait ratings, with those high in Restraint correcting for their automatic prejudices in the revised trait impression task in the individual (but not category) target construal condition. However, once again, I found no interaction of Concern with automatically-activated attitudes in predicting impressions in the category (or individual) target construal condition. Just as before, Concern was positively correlated with trait ratings and pro-Black trait preferences regardless of participant automatic prejudice level (Zabel & Olson, 2014). Thus, findings from two studies indicate that high Restraint individuals tend to correct for their automatic biases in individual target construal contexts. High Concern individuals tend to utilize a “like everyone” (and “if you must choose, favor Blacks”) approach but fail to proportionately correct for their automatic biases in their subsequent judgments and behaviors. A critical research question then becomes *why* does this failure to correct for automatic biases exist among high Concern individuals? These findings (Zabel & Olson, 2014) raise the intriguing possibility that the very individuals who perceive

themselves as egalitarian and fair-minded may be the least likely to introspect and correct for their own automatic racial biases.

Theoretical Perspectives Underlying a Lack of Bias Correction

There are at least two potential explanations as to why high Concern individuals may fail to correct for their automatic racial attitudes. I aim to test these alternative hypotheses in the proposed research: 1) High Concern individuals may lack the skill and experience to introspect (i.e., consciously look inward) on their automatic racial biases, and 2) High Concern individuals may find introspecting on their racial biases to be a threat to their egalitarian self-image. Each of these possibilities may influence high Concern Whites to employ “like everyone” and “prefer Black targets” heuristics in impression formation. These strategies, while leading to pro-Black judgments and behaviors, fail to account for the strength of one’s automatic racial attitudes in impression formation. The proposed research allows examination of the degree to which these potential explanations may be driving failure to introspect, and ultimately, failure to correct for automatic prejudices among egalitarian-oriented (high Concern) individuals. Next, I highlight research on the process of introspection, given its predicted relevance to bias correction failures among high Concern individuals. Then, I more specifically hone in on why introspection failures may be occurring, and how the tendency for high Concern Whites to introspect and correct for their automatic racial prejudices may be increased.

Introspection

Introspection, or the conscious act of looking inward and “thinking about one’s thoughts and feelings” (Wilson et al., 1993, p. 331) has been a topic of interest among philosophers and psychologists for centuries (Hunt, 2007). Philosophers as early as Socrates were interested in how individuals acquire self-knowledge through the process of introspection. David Hume, in his “A Treatise of Human Nature” questions the capacity or ability of individuals to accurately introspect. Wilhelm Wundt severely questioned introspection as a proper means to inferring knowledge about human mental states and the self in an empirical, replicable manner (Blumenthal, 1975; Hunt, 2007). Although admitting introspection could be fallible, William James contended that it could be a valuable tool in experimental methods regarding more conscious mental processes, stating that “introspective observation is what individuals have to rely on first and foremost and always” (James, 1890, pp. 185). Of course, prominent psychologists (i.e., Sigmund Freud, John Watson) have pondered the structure, validity, utilization, and accessibility of introspective processes in psychology, albeit from very different theoretical orientations (Hunt, 2007).

Prior research differentiates between two forms of introspection: descriptive and explanatory (Sedikides, Horton, & Gregg, 2007). Descriptive introspection focuses on the “what” components of self-knowledge; that is, the contents of self-knowledge regarding a particular attitude, aspect of the self-concept, feeling, or behavior. Explanatory introspection focuses on the “why” components of self-knowledge; that is, reflecting on reasons why one has a certain attitude, feeling, or behaves a certain way (Wilson, Dunn,

Kraft, & Lisle, 1989; Wilson et al., 1993). Explanatory introspection tends to lead to more inaccurate self-knowledge than descriptive introspection, given that it may be more prone to rationalization and availability cues (Hixon & Swann, 1993; Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984). For instance, focusing on “what” their feelings actually were (as opposed to “why” they felt a particular way) increased correspondence between peoples’ self-reported affect and actual behavior (Carver & Scheier, 1981; Fazio, Chen, McDonel, & Sherman, 1982), as well as implicit and explicit attitude measures (Gawronski & LeBel, 2008).

Introspection Versus Awareness

To my knowledge, no research to date has examined the precise role that active *introspection* of automatic racial attitudes may play in bias correction. One recent theoretical perspective has begun to zero in on the important roles that both *awareness* and adjustment (motivation) may hold in bias correction. Specifically, Hofmann, Gschwendner, Nosek, and Schmitt (2005) proposed a moderated consistency model for implicit and explicit attitudes, in which awareness “affects how strongly the conscious representation of an attitude reflects the implicit attitude, whereas adjustment affects how strongly the overt expression of an attitude depends on the preceding propositional representation” (Hofmann et al., 2005, pp. 28). From this theoretical orientation, awareness and adjustment are posited to interact to predict the correspondence between implicit attitudes and behaviors and judgments. One recent research study from this perspective indicates that a personality orientation toward self-awareness increases bias correction for automatic prejudices among individuals high in motivation to control

prejudiced reactions (Hofmann et al., 2005). Another recent research study indicates that a personality orientation toward self-awareness and a motivation to introspect may increase implicit prejudice-explicit measure correspondence (Gschwendner, Hofmann, & Schmitt, 2006). However, it is important to note that in neither of these studies was the active process of introspecting on automatic racial biases and subsequent impression formation repercussions examined, nor were the bias correction tendencies of egalitarian-oriented (high Concern) individuals examined, which is central to the proposed work.

From a Moderated Consistency Model perspective, awareness is posited to *increase* the correspondence between implicit and explicit attitudes, whereas adjustment (motivation) is posited to *reduce* implicit-explicit correspondence. However, it is important to note that awareness is distinct from introspection. For instance, awareness may be activated when individuals are primed with race (as in an Implicit Association task; Monteith, Voils, & Ashburn-Nardo, 2001; Plant et al., 2010; Strack & Deutsch, 2014), but awareness does not necessarily lead to active introspection. Indeed, I contend that once awareness occurs, it is the nature of the adjustment motive (i.e., accuracy, motivation to control prejudiced reactions) that in turn either activates or fails to activate introspection. Although previous research indicates that motivation can enhance implicit-explicit correspondence (Gschwendner et al., 2006; Phillips & Olson, 2014), I contend that the effects of awareness on implicit attitude-explicit measure correspondence are entirely dependent on the underlying nature of situationally-activated motives. For instance, activation of honesty or accuracy motivation may evoke greater introspection of automatic racial attitudes and implicit attitude-explicit measure correspondence (Phillips

& Olson, 2014; see also Gschwendner et al.'s 2006), whereas activation of conflict-avoidance related motives may also evoke greater introspection of automatic racial attitudes but implicit attitude-explicit measure *divergence*. Furthermore, as the present study examines, activation of some prejudice-related motives (i.e., egalitarian-oriented motives) may squelch introspection of automatic racial attitudes among high Concern individuals (Zabel & Olson, 2014) in race-related contexts, leading to a failure to correct for automatic prejudices.

Lacking the Skill to Introspect

In returning and focusing the lens to introspection, recall that one cognitive-based possibility as to why egalitarian-oriented (high Concern) Whites may fail to introspect on their automatic racial biases is because they lack the skill and experience to do so. Indeed, much research indicates that introspection is fallible and can actually be an inaccurate source of self-knowledge (Nisbett and Wilson, 1977). Another point is that human beings do not introspect nearly as much as our naïve theories may suggest (Csikszentmihalyi & Figurski, 1982), suggesting that individuals generally may lack experience in introspection. Furthermore, introspection can actually hurt or damage self-knowledge (Hodges & Wilson, 1993; Wilson, Kraft, & Dunn, 1989). Reasons-generated attitude change (Hodges & Wilson, 1993) argues that the process of looking within regarding one's feelings or actions is prone to availability heuristic-related (Tversky & Kahneman, 1973) pitfalls that influence an overreliance on easily verbalized, albeit incorrect reasons as a source of self-knowledge.

In connecting introspection to dual process theories regarding the interplay of automatic and more controlled processes, automatic processes are argued to be less available to introspection (Wilson, 2009; Wilson & Dunn, 2004). Nevertheless, Wilson (2009) contends that “self-knowledge is less a matter of careful introspection than of becoming an excellent observer of oneself and deducing the nature of one’s nonconscious dispositions and preferences” (pp. 386). Thus, although Wilson (2009) seems to cast doubt that conscious (and accurate) introspection as to one’s automatic attitudes is probable, he does concede that self-knowledge as to automatic attitudes may be possible via implicitly inferring attitudes based on one’s behaviors and reactions. This latter point is congruent with previous empirical research from a self-perception theory perspective (Bem, 1967; Strack, Martin, & Stepper, 1988; Wells & Petty, 1980). Wilson’s (2009) research and the aforementioned research (Hodges & Wilson, 1993; Nisbett and Wilson, 1977; Wilson et al., 1989) suggest that individuals may lack skill in introspection and may have a reduced capacity to accurately introspect on more automatic psychological processes.

Importantly, egalitarian-oriented (high Concern) individuals may especially lack skill in introspecting on their automatic racial biases, relative to conflict avoidance-oriented (high Restraint) individuals, perhaps in large part due to a lack of experience. Indeed, it would be essential for high Restraint Whites to introspect on their automatic prejudices more frequently to avoid bi-directional sources of conflict associated with both appearing prejudiced *and* bending over backward toward Blacks. High Concern individuals, in having the unidirectional goal of avoiding prejudice, would not need to

introspect on their automatic racial attitudes in order to fulfill activated Concern goals. In the present research, I examine the degree to which lacking the experience and skill to introspect may be driving high Concern individuals' tendencies to avoid correction for automatic prejudices in impression formation.

Motivational Influences on Introspection

Another motivationally-based possibility is that high Concern individuals do have the capacity to accurately introspect on their automatic racial biases, but avoid doing so to preserve their egalitarian self-image. Of course, fundamental to a motivational perspective is that high Concern individuals do have the capacity and skill to accurately introspect on their automatic racial biases, a tenet congruent with recent research. Specifically, individuals can accurately introspect on their automatic racial attitudes under certain *motivated* (i.e., motivation for accuracy) conditions (Phillips & Olson, 2014). I contend that high Concern individuals may be motivated to avoid introspection, given that doing so would be a threat to their egalitarian self-image. This avoidance, in turn, leads to a failure to correct for automatic biases.

In contrast, for high Restraint Whites, introspecting on automatic racial attitudes is likely not threatening to their self-image, given that their motive to avoid prejudice is more externally-driven. Such reasoning is consistent with findings that high Restraint individuals correct for their automatic prejudices in trait judgments, whereas high Concern individuals fail to correct proportionately for their automatic prejudices (Zabel & Olson, 2014). Of course, this perspective implies that the process of introspecting on automatic racial biases should increase bias correction among high Concern Whites. In

what follows, I lay out theoretical perspectives supporting the notion that introspecting on automatic racial biases may threaten the self-image of high Concern Whites that in turn squelches bias correction, before moving on to examining possibilities for reducing the self-image threat of introspection.

Several theoretical positions support the possibility that a threat to the self-image of high Concern individuals is the primary mechanism driving failure to introspect and subsequently, a failure to correct for automatic biases. For instance, self-discrepancy theory (Higgins, 1987) states that a discrepancy between one's ought (i.e., "I should treat everyone the same") and actual self (i.e., "I am prejudiced at some level") is threatening to one's self-image and can lead to anxiety (Higgins, 1987). Thus, avoiding introspection on one's racial attitudes may be a strategy high Concern individuals use to avoid a clash between actual and ought selves. Relatedly, self-awareness theory (Duval & Wicklund, 1972) argues that increased self-focused attention (i.e., introspection) leads to a greater likelihood of experiencing discrepancies between one's ought and actual self that might lead to anxiety. Cognitive dissonance theory (Festinger, 1957) argues that a context in which one has two related but incompatible thoughts is an anxiety-producing situation that one is motivated to escape. Self-consistency theory (Aronson, 1969; 1999) extends cognitive dissonance theory by arguing that dissident cognitions are anxiety-producing and aversive to the extent to which they contradict a person's self-concept. These theories share in common that a mismatch between thoughts or behaviors that have implications for one's self-image is an aversive, threatening experience that one is motivated to avoid.

Another theoretical model that emphasizes the potential threat of appearing prejudiced to those with implicitly-prejudiced but explicitly unprejudiced attitudes and egalitarian-activated motives is the Self-Regulation of Prejudiced Responses Model (Monteith, 1993). This model attests that stereotypes are automatically-activated among perception of a Black target that may influence White individuals to behave in a discrepant way from their egalitarian motives. This experience is posited to be aversive and laden with negative affect (e.g., guilt) for White individuals, leading to the development of cues for control that help reduce such aversive behaviors. Individuals for whom egalitarian concepts are chronically-activated have even demonstrated reduced stereotype activation as a self-regulatory strategy (Moskowitz, Gollwitzer, Wasel, & Schall, 1999). Fundamental to the Self-Regulation of Prejudiced Responses Model is that a discrepancy between behavior and explicit racial attitudes or egalitarian motives is a threatening, uncomfortable experience that individuals are motivated to avoid (see Monteith & Mark, 2005 for a review).

Perhaps most directly, Aversive Racism Theory may shine light on why egalitarian-oriented (high Concern), but automatically prejudiced individuals are failing to introspect on their automatic prejudices (Gaertner & Dovidio, 1986). The theory argues that although aversive racists seek to maintain an egalitarian self-image and self-report unprejudiced attitudes toward Blacks, such individuals still harbor negative biases that can be revealed by implicit measures. Thus, according to Aversive Racism theory, an aversive racist is characterized by positive, self-reported racial attitudes and an egalitarian self-concept, but more negative, prejudicial automatic racial attitudes (Dovidio &

Gaertner, 2004). In other words, high Concern and automatic prejudice defines such individuals (Song Hing et al., 2008).

Situations in which prejudice concerns are salient and inescapable are theorized to be especially threatening to aversive racists, as such situations may force aversive racists to face the conflict between their egalitarian self-image and their automatic racial prejudices, increasing the likelihood of automatic prejudice unwittingly influencing their behaviors. Indeed, this latter viewpoint is buttressed by interracial interaction research indicating that automatic prejudice can “leak” into one’s nonverbal behavior (Dovidio, Kawakami, & Gaertner, 2002), which in turn can have negative consequences for an interaction (Fazio et al., 1995; McConnell & Leibold, 2001; but see Shelton, Richeson, Salvatore, & Trawalter, 2005).

Fundamental to Aversive Racism Theory is that White individuals engage in specific strategies to maintain an egalitarian self-image in contexts in which automatic prejudices are often unacknowledged, but that automatic prejudices may still lead to discriminatory actions in insidious ways (Gaertner & Dovidio, 1986). For instance, aversive racists avoid race-related contexts which may force them to confront their egalitarian self-image and underlying automatic prejudices (Gaertner, 1973). However, when enough contextual information or ambiguity exists for a non-race based attribution to account for favoring White targets (e.g., less normative concerns and additional target information), the automatic prejudices of aversive racists tend to unwittingly influence behaviors and judgments (i.e., favoring White relative to Black targets; Dovidio & Gaertner, 2000; Gaertner, 1973; Hodson, Dovidio, & Gaertner, 2002; Song Hing et al.,

2008). Indeed, with increasing contextual information and attributional ambiguities, favoring Whites is no longer a self-image threat to aversive racists. Aversive racists tend to avoid favoring Whites only when norms are clear (Dovidio & Gaertner, 2000; Gaertner, 1973; Hodson et al., 2002; Song Hing et al., 2008; see Pearson, Dovidio, & Gaertner, 2009 for a review).

Recall that my research thus far indicates that high Concern individuals tend to employ a “like everyone” approach in impression formation (Zabel & Olson, 2014) and prefer Blacks when forced to choose between Black and White targets (Zabel & Olson, 2014). In both, cases, however, high Concern individuals do not correct proportionately for their automatic racial attitudes in trait judgments. This is in stark contrast to more conflict avoidance oriented (high Restraint) individuals, who correct for their automatic prejudices. Thus far, I have highlighted theoretical support for the notion that a mismatch between automatic prejudices and egalitarian-motives is a self-image threat to high Concern White individuals that is aversive and anxiety-producing. I contend that failing to introspect on one’s automatic racial attitudes is a strategy that protects the self-concept of high Concern individuals from the threat of a mismatch between their egalitarian self-image and prejudiced automatic racial attitudes.

Once forced to introspect on their automatic racial biases, and unable to avoid the self-image threat, I expect high Concern Whites to correct for their automatic racial attitudes in a manner congruent with previous research (Dunton & Fazio, 1997; Olson & Fazio, 2004; Zabel & Olson, 2014). Indeed, once a self-image threat cannot be avoided, automatic bias corrections (Dunton & Fazio, 1997; Fazio et al., 1995; Olson & Fazio,

2004; Wegener & Petty, 1997) become the optimal strategy for high Concern individuals in order to “save face” and preserve their egalitarian self-image. Indeed, this reasoning is congruent with previous research in which egalitarian-oriented Whites only expressed interest in a prejudice correction program under instances of greater race saliency or awareness (Plant et al., 2010). Indeed, only in instances when a self-image threat cannot be avoided would greater and accurate racial attitude introspection be most functional for egalitarian-oriented individuals. A motivation mechanism perspective provides an alternative to a more cognitive (i.e., lack of skill) perspective in understanding why egalitarian-oriented (high Concern) individuals may fail to introspect on their automatic racial biases. Furthermore, each provides alternative explanations for the findings of previous research (Zabel & Olson, 2014).

Although beneficial to protecting egalitarian-oriented White individuals’ self-images, failing to introspect on automatic biases could ironically lead egalitarian-minded Whites to engage in behaviors or judgments that proportionately do not correct for automatic attitudes. This is an important distinction to examine, as the very people most concerned about redressing past inequalities and fair treatment may be the most unlikely to correct for their own automatic racial biases due to the potency of self-image threat. If this is found to be the case in the present research, a crucial research question then becomes *how* can the tendency for high Concern White individuals to introspect on their automatic racial biases be increased? Increasing introspection would be crucial to leading to more appropriate corrections for automatic prejudice that may have significant positive intergroup consequences. One potential mechanism to reduce the threat of introspecting

on racial biases for high Concern individuals may be self-affirmations, which I turn to next.

Self-Affirmation Theory

Self-affirmation theory (Steele, 1988) contends that individuals engage in a variety of strategies to affirm the self when a threat to one's self-image is derived from either meta-perception of others or one's own behavior. According to self-affirmation theory, the purpose of the self is to "maintain a perception of global integrity and overall moral and adaptive adequacy" (Steele, Spencer, & Lynch, 1993, p. 885). Self-affirmations not only reaffirm the threatened self, but also preemptively protect against future threats (Steele & Liu, 1983; Steele et al., 1997; Tesser & Cornell, 1991). Moreover, self-affirmations are flexible in protecting the self in that they do not have to be domain-related to the nature of the threat itself. For instance, in one study, self-affirmations (i.e., writing about one's positive features) led smokers to be more open to self-threatening graphic smoking images that demonstrated the dangers of smoking relative to those in a non-self-affirmation control condition (Harris, Mayle, Mabbott, & Napper, 2007; see also Armitage, Harris, Hepton, & Napper, 2008). Other research indicated that self-affirmations can buffer against physiological stress responses to presenting an impromptu speech (Kirschbaum, Pirke, & Hellhammer, 1993) and to stressful midterm examinations (Sherman, Bunyan, Creswell, & Jaremka, 2009). Indeed, the ability of self-affirmations to buffer against a broad range of psychological threats is conceptualized in its inclusion as part of the psychological immune system (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998).

Other research indicates that a self-affirmation can preemptively reduce the threat of a self-threatening upward social comparison, as well as reduce the implications of a social comparison threat after experience of that threat (Tesser & Cornell, 1991). Self-affirmations buffer against the tendency for White individuals to derogate out-group members as a self-affirmation strategy (Fein & Spencer, 1997). Moreover, evidence suggests that individuals are more likely to pay attention to self-disconfirming and incongruent information following a self-affirmation, suggesting that confirmation bias tendencies may be reduced following a self-affirmation (Cohen, Aronson, & Steele, 2000). In short, in situations in which an individual's self-integrity is threatened in any manner, affirming any aspect of the self can serve as a preemptive tool or solution to mitigate the effect of that threat on behavior.

In refocusing the lens to my core research question, it is likely that a self-affirmation could buffer against the threat of introspection to the self-images of high Concern Whites. A self-affirmation would allow high Concern individuals the “self-credits” (Monin & Miller, 2001) to introspect on their automatic racial biases, which otherwise would be an overly threatening and aversive experience. I contend this introspection, in turn, may be essential in leading to corrections for automatic prejudice among high Concern Whites. In the present study, I examine whether a self-affirmation bolsters the self-integrity of egalitarian-oriented (high Concern) individuals sufficiently to mitigate the threat of introspecting on their automatic biases.

Self-affirmations appear to be most effective in reducing the potential *threat* of stimuli or contexts (Sherman & Hartson, 2011). Thus, if self-affirmations enhance the

degree to which high Concern individuals correct for their automatic biases under conditions of introspection, it is likely that introspection's bias correcting effect is being driven by reducing self-image threat. Additionally, the results of other studies (Zabel & Olson, 2014) showing a lack of automatic prejudice correction among egalitarian-oriented (high Concern) could then be explained by a failure to introspect due to a self-image threat.

However, if self-affirmations fail to moderate the expected effect of introspection on automatic prejudice correction among egalitarian-oriented (high Concern) individuals, then it will become less likely that a motivationally-driven mechanism to avoid self-image threat is driving introspection failure among high Concern Whites. In this case, it would become more likely that mechanistic explanations are more cognitive in nature, such as lacking the skill to accurately introspect.

Chapter 2: Hypotheses

Hypotheses

In the current research, I measured Concern and Restraint motivation to control prejudiced reactions (Dunton & Fazio, 1997), as well as automatically-activated racial attitudes (Fazio et al., 1995). Subsequently, participants were randomly assigned to engage in a self-affirming activity (or not), as well as to either introspect on their automatic racial attitudes (or not) in a counterbalanced fashion. In all conditions, race-related awareness was held consistent. All participants next completed an impression formation task of Black and White group and individual targets (Olson & Fazio, 2004). Relevant potential individual difference moderators (e.g., Private and Public Self-Consciousness: Fenigstein, Scheier, & Buss, 1974), as well as explicit measures of prejudice (e.g., Symbolic Racism: Henry & Sears, 2002) were also assessed.

Replicating previous research, I expected high Concern (but not Restraint) individuals to provide more positive trait ratings toward both Black and White individual and group targets. However, I expect that egalitarian-oriented (high Concern) individuals do have access to the content of their automatic racial attitudes and are subsequently able to correct for the influence of their automatic prejudices under the right conditions. Specifically, in the current research, I expected that correction for automatic prejudice would be greatest among high Concern individuals who first were self-affirmed and then introspected on their racial biases. Precisely, I expected self-affirmations to reduce the self-image threat of introspecting on racial prejudices among high Concern Whites. I expected this in turn to lead to greater and more accurate racial attitude introspection

among high Concern Whites, congruent with previous research indicating that individuals do have access to the content of their automatic racial attitudes (Phillips & Olson, 2014). Once introspection as to one's racial attitudes occurred, I expected correction for automatic prejudices to become the default strategy among high Concern Whites, in patterns analogous to those verified in previous research (Fazio & Olson, 2014). Given Restraint motivation's conflict avoidance focus, I expected that it would not interact with self-affirmation or racial attitudes introspection in predicting corrections for automatic prejudice.

Chapter 3: Method

Participants

A total of 228 (106 female and 122 male) White undergraduates completed the experiment for credit in their psychology courses.

Materials and Procedure

Motivation to control prejudiced reactions (MCPR). Upon signing up for the primary study and prior to their arrival to the laboratory, participants completed an online version of Dunton and Fazio's (1997) 17-item Motivation to Control Prejudiced Reactions (MCPR) questionnaire to assess Concern and Restraint motivations. Participants rated the degree to which they agreed with a variety of statements such as, "If I were participating in a class discussion and a Black student expressed an opinion with which I disagreed, I would be hesitant to express my own viewpoint" (Restraint) and "I feel guilty when I have a negative thought or feeling about a Black person" (Concern) using a 1 (*strongly disagree*) to 5 (*strongly agree*) response range. In the current study, $\alpha = 0.82$. Descriptive statistics for all variables are located in Table 1. Furthermore, descriptive statistics for all variables split by experimental condition are located in Tables 2-6. All tables and figures are located in the appendices.

Upon first arriving to the laboratory, participants were seated at individual cubicles containing computer monitors at which they completed all subsequent tasks. Participants were told the experiment was about first impressions and would also involve categorizing visual stimuli. After consenting to participate, participants first completed two individual difference measures relevant to the primary hypotheses.

Private self-consciousness. First, participants completed Fenigstein et al.'s (1975) 10-item private self-consciousness scale as an individual difference level assessment of general introspective tendencies and self-awareness. Participants rated the degree to which a variety of statements, such as "I'm always trying to figure myself out" and "I reflect about myself a lot" were characteristic of them using a 0 (*extremely uncharacteristic*) to 4 (*extremely characteristic*) response range. In the current study, $\alpha = 0.49$.

Public self-consciousness. Participants next completed Fenigstein et al.'s (1975) 7-item public self-consciousness scale as an individual difference level assessment of general social self-awareness. Participants rated the degree to which a variety of statements, such as "I get embarrassed very easily" and "Large groups make me nervous" were characteristic of them using a 0 (*extremely uncharacteristic*) to 4 (*extremely characteristic*) response range. In the current study, $\alpha = 0.78$.

Automatic prejudice. Following completion of these two individual difference measures, participants completed the core measures and experimental manipulations. First, participants completed Fazio et al.'s (1995) unobtrusive priming measure of automatically-activated racial attitudes. Block 1 was comprised of practice trials, where each of sixteen adjectives was presented to participants, which they categorized as either positive or negative. Blocks 2 through Block 5 contained the critical trials. Participants were told that this part of the experiment concerned their ability to perform two tasks simultaneously. On a given trial, a face prime was presented prior to each adjective. Participants were instructed to pay attention to the faces because they would be asked

questions about them later, but only to respond to the valence of each target adjective.

Participants were instructed to complete the task as quickly and accurately as possible.

The primes were 24 photos of faces of various races. Of the faces, 8 were White, 8 were Black, and 8 were fillers of White and other-raced faces to obscure my interest in race. The primes were gender-balanced, such that there was an equivalent number of female and male Black, White, and other-race primes. On each trial, a prime was presented for 315 ms, followed by a 135 ms interval, followed by the target adjective (the same 16 target adjectives used in Block 1). Thirty-two of the 48 trials presented in Blocks 2-5 consisted of critical pairs of photos matched by target adjective and sex but varying by race. Over the course of Blocks 2-5, each member of each Black and White photo prime pair was presented four times with the same positive and negative target adjectives, resulting in a total of 128 critical trials. Participants' responses and latencies in categorizing the target adjectives as a function of race prime (Black vs. White) were recorded on each trial for later use in computing automatically-activated racial attitudes.

Self-affirmation manipulation. Participants were randomly assigned to either complete a frequently-used (Adams, Tormala, & O'Brien, 2006; Fein & Spencer, 1997; Sherman, Nelson, & Steele, 2000) self-affirmation (SA) manipulation or not. Participants assigned to the SA condition ranked a list of 11 provided topics in terms of personal importance. Following, they described their most important value and the reason for its singular importance, as well as provided an example from their life that demonstrated the importance of that value to them.

Introspection manipulation. First, I utilized a prompt to minimize race-related awareness differences between racial attitude introspection conditions. Specifically, all participants in all conditions read the following instructions on their computers: “Part of the importance of participating in psychology research at the University of Tennessee is to complement and extend your classroom learning. Next, you will read about a research finding that has been well-supported in previous research to enhance the value of this experiment to your learning, prior to moving on to the next part of the experiment.” All participants next read the following statement on their computer screens “Automatic prejudices (negative feelings toward Black Americans) are pervasive in American society. Most Americans are prejudiced toward Black Americans, even if they are not aware of this bias. This is an important finding in contemporary psychology, and thus relevant to bring to your attention today.” This prompt rules out racial awareness differences among the racial attitude introspection conditions.

Participants were then randomly assigned to either undergo a racial attitudes introspection manipulation or not. Participants in the introspection condition read and responded to the following statements: “In the next part of the experiment, please take a few minutes to reflect on your initial “gut feelings” and racial attitudes (feelings) toward Black Americans. Think about what your “gut feelings” and racial attitudes are toward Black Americans, and take a few minutes to write about these feelings. This task will help organize your thoughts for tasks later in the experiment.” In comparison, control condition participants did not engage in introspection. The introspection manipulation was formulated loosely based on previous research (Sedikides et al., 2007). Participants

completed the introspection and self-affirmation manipulations in a counter-balanced fashion.

Impression formation task. Following the self-affirmation and introspection manipulations, participants completed a trait rating task in which they formed impressions of 25 specific individuals and 25 specific groups depicted in various occupational and everyday settings. Regarding the individual photos, nine individuals were fillers of White, Arab, and Latino individuals included to obscure the nature of the task (no mention of our interest in race was made). The remaining 16 individual photos consisted of eight critical Black and eight critical White individuals previously matched by sex, as well as status and independence of their occupation (Fazio & Dunton, 1997). Female pairs consisted of a Black receptionist matched with a White business woman, a Black nurse matched with a White pharmacist, a Black repair woman matched with a White painter, and a Black cashier matched with a White police officer. Male pairs consisted of a Black minister matched with a White professor, a Black businessman matched with a White architect, a Black gardener matched with a White sanitation worker, and a Black potter matched with a White brick layer. These photos have been utilized several times in previous research as a dependent variable measure of pro-Black trait ratings (Olson & Fazio, 2004; Zabel & Olson, 2014).

Participants also formed impressions of 25 groups, nine of which were fillers of White, Arab, and Latino individuals. The remaining 16 group photos consisted of eight critical Black and eight critical White groups previously pilot-tested to be equivalent on dimensions such as likeability and matched by occupation (e.g., sanitation worker) or

general characteristics (e.g., college students). Critical female pairs consisted of Black and White bridal parties, business women, conference goers, and college groups. Critical male pairs consisted of Black and White lawyers, firefighters, sanitation workers, and basketball teams.

For each individual or group target photo, participants rated the degree to which each individual or group possessed each of five desirable traits (i.e., intelligent, industrious, likeable, honest, reliable) using a 0 (*not at all*) to 6 (*very much so*) scale. Presentation order of the targets was randomized, but participants always completed the five trait ratings in a fixed order above before moving on to the next target. Furthermore, the presentation order of whether participants completed all the individual or group target ratings first was counterbalanced. Reaction times for each participant's judgments were recorded.

Pro-black/anti-black attitudes. Following the impression formation task, participants completed Katz and Hass's (1988) 10-item pro-Black and 10-item anti-Black attitude questionnaire. Pro-Black attitude questions consisted of items such as "Many Whites show a real lack of understanding of the problems that Blacks face" and "Blacks have more to offer than they have been allowed to show," and anti-black attitude questions consisted of items such as "On the whole, Black people don't stress education and training," and "One of the biggest problems for a lot of Blacks is their lack of self-respect." Participants rated the degree to which they agreed with each statement using a 1 (*strongly disagree*) to 7 (*strongly agree*) response range. In the current study, internal

reliabilities for the pro-Black and anti-Black attitude questionnaires were 0.56 and 0.67, respectively.

Subtle and blatant prejudice. Subsequently, participants completed Pettigrew and Meertens's (1995) 8-item subtle prejudice scale and 8-item blatant prejudice scale. Participants rated the degree to which they agreed with each of the subtle (e.g., "Black Americans here should not push themselves where they are not wanted") and blatant (e.g., "White and Black Americans can never be really comfortable with each other, even if they are close friends") prejudice items using, for the most part, a 1 (*strongly disagree*) to 5 (*strongly agree*) response range. However, some of the scale items consisted of varying response ranges (e.g., four response options instead of five). Thus, the z scores of all items were calculated prior to computing reliabilities and individual mean scores on each scale. In the current study, internal reliabilities for the subtle and blatant prejudice scales were 0.80 and 0.76, respectively.

Symbolic racism. Following, participants completed the symbolic racism 2000 scale (Henry & Sears, 2002). Participants rated the degree to which they agreed with statements such as "It's really just a matter of some people not trying hard enough; if Blacks would only try harder they could be just as well off as Whites," and "Over the past few years, Blacks have gotten less than they deserve" (reverse-coded) using a 1 (*strongly disagree*) to 4 (*strongly agree*) response range. However, one scale item consisted of only three response options (relative to four). Thus, the z scores of all items were calculated, prior to computing the scale reliability and individual mean score. In the current study, $\alpha = 0.70$. Following, participants completed demographic items assessing

their sex and race. Furthermore, they completed an open response item, in which they responded to the following statement: “What are your general feelings about this experiment? Do you have any thoughts about what research questions this experiment is examining?” Subsequently, participants were debriefed about the true nature of the experiment and dismissed.

Self-affirmation manipulation coding. Following completion of the study, three White, female laboratory research assistants individually rated the extent to which participants followed the self-affirmation directions in the experimental (affirmation) condition using a 0 (*not at all*) to 4 (*extremely well*) response range. They also rated how affirmed they thought the participant would be after writing their statement using a 0 (*not at all*) to 4 (*extremely*) response range. Additionally, the number of words written in each self-affirmation statement was recorded.

Introspection manipulation coding. The same laboratory research assistants also rated the extent to which participants followed the introspection directions in the experimental (introspection) condition using a 0 (*not at all*) to 4 (*extremely well*) response range. They also reported how negatively-biased participants admitted to being against African-Americans using a 0 (*not at all*) to 4 (*extremely*) response range. As an additional measure, they reported, based on their own general and more subjective impressions, how racially biased the participant seemed to be using a 0 (*not at all*) to 4 (*extremely*) response range. Additionally, the number of words written in each introspection statement was recorded.

Chapter 4: Results

Data Preparation

Motivation to control prejudiced reactions. Replicating previous research (Dunton & Fazio, 1997, Olson & Fazio, 2004), concern and restraint motivational factors emerged in principal component analysis using varimax rotation. I used the factor score coefficients to compute the two factor scores for participants. One participant scored greater than three standard deviations from the mean (i.e., $z = -3.47$) in Concern motivation, and thus was removed from primary analyses.

Automatic prejudice. One participant was removed from primary analyses for having error rates of greater than 30% on critical trials (i.e., 35.20% of trials). Consistent with previous research (Greenwald, Nosek, & Banaji, 2003), one participant was removed from primary analyses for taking longer than 2000 ms to categorize target words on more than 10% of trials (i.e., 68.00% of trials). Critical trials of which responses were incorrect (2.65% of cases) or in which raw latencies were less than 300 (<.01% of cases) or greater than 2000 ms (<.01% of cases) were removed prior to aggregation. Then, all response latencies were reciprocally-transformed to better approximate normality. Using reciprocally-transformed response latencies, I computed participants' mean response latency difference scores on critical trials in which a Black prime preceded a negative, relative to a positive adjective, as well as on critical trials in which a White prime preceded a negative, relative to a positive adjective. Next, we subtracted participants' scores on the latter difference score from their scores on the former difference score to derive an automatically-activated racial attitude estimate used in primary analyses.

Positive scores indicated pro-White automatically-activated attitudes. Automatic attitude estimates ranged from $-.00035$ to $.00043$ ($M = .00004$, $SD = .00013$; raw $M = 13.74$, raw $SD = 68.48$), were positively-distributed (144 of the 228 scores were positive), and differed from 0, $t(227) = 4.37$, $p < .001$, indicating a tendency for pro-White automatically-activated attitudes. One participant had an automatic attitude estimate greater than three standard deviations from the mean (i.e., $z = 3.02$), and was removed from primary analyses.

Trait ratings. The trait ratings were highly correlated within critical individual and group targets. Reliability analyses revealed coefficient alphas ranging from $.72$ to $.88$ across critical individual target photos ($\alpha_M = .83$), and from $.81$ to $.90$ across critical group photos ($\alpha_M = .85$). Thus, I combined the five traits into a single mean for each target across the critical individual and group targets. I then computed difference scores indicative of Pro-Black individual and group trait ratings for the matched pairs. Specifically, I subtracted participants' mean individual trait ratings of critical White targets from their mean individual trait ratings of critical Black targets (separately for both individuals and groups). Positive difference scores represented more favorable trait ratings of Black, relative to White targets. Pro-Black individual trait rating difference scores ranged from -1.58 to 1.33 ($M = .07$, $SD = .38$), whereas Pro-Black group trait rating difference scores ranged from $-.55$ to 2.20 ($M = .45$, $SD = .40$). The mean of individual Pro-Black trait rating difference scores differed from 0, $t(227) = 2.95$, $p < .01$, as did the mean of group Pro-Black group rating difference scores, $t(203) = 16.23$, $p < .001$, indicating a tendency for participants to indicate more favorable trait ratings of

Black individual and group targets relative to White targets. This difference is consistent with previous research showing more positive reported impressions of Blacks relative to Whites (e.g., Biernat & Vescio, 1993). A paired-samples *t*-test indicated that pro-Black ratings toward group targets were more favorable than those toward individual targets, $t(203) = -13.43, p < .001$. Additionally, given the positive correlation between pro-Black group and individual ratings ($r = .49, p < .001$), I created a variable that collapsed across group and individual targets in which I subtracted participants' mean trait ratings toward both White group and individual targets from participants' mean trait ratings toward both Black group and individual targets. These Pro-Black difference scores ranged from -.88 to 1.40 ($M = .24, SD = .34$), and were significantly different from 0, $t(227) = 10.55, p < .001$.

Four participants' pro-Black difference scores collapsed across group and individual targets were greater than three standard deviations from the mean (i.e., z 's = 3.00, -3.22, -3.26, & 3.42), and these participants were omitted from primary analyses. Additionally, examination of free responses at the end of the experiment uncovered that four participants were precisely aware of either the true purpose of the study or how automatically-activated attitudes were assessed within the study. Due to concerns regarding demand characteristics, these participants were omitted from primary analyses. Finally, a thorough look at the data set uncovered one participant who responded almost exclusively to the trait rating task with the rating of "7" (i.e., toward 82.4% of targets). Furthermore, this participant frequently made judgments in less than 300 ms, suggesting disengagement within the trait rating task. Thus, this participant was removed from

primary analyses. Removing the aforementioned outliers did not alter the patterns of the effects highlighted in the following primary analyses.

Self-affirmation manipulation coding. Self-affirmation word count ($M = 74.12$, $SD = 41.96$) was conducted by a trained research assistant, whereas three research assistants rated the degree to which each participant followed directions and the degree to which each participant self-affirmed based on their written statement. Research assistant ratings on the degree to which participants followed directions were highly correlated (all r 's $> .59$, p 's $< .001$, $M_r = .67$). Research assistant ratings on the degree to which each participant self-affirmed based on their written statement were also highly correlated (all r 's $> .54$, p 's $< .001$, $M_r = .61$). Thus, the composite of the ratings provided by the three research assistants for each rating among each participant were averaged for use in primary analyses.

Introspection manipulation coding. Introspection word count ($M = 58.19$, $SD = 36.40$) was conducted by a trained research assistant, whereas three research assistants rated the degree to which each participant followed directions, admitted to being negatively biased against African-Americans, and appeared to be racially biased based on their own subjective impressions. Correlations between research assistant ratings on the degree to which each participant followed directions (all r 's $> .44$, p 's $< .001$, $M_r = .45$), admitted to being negatively-biased toward African-Americans (all r 's $> .67$, p 's $< .001$, $M_r = .70$), and appeared to be racially biased based on their own subjective impressions (all r 's $> .54$, p 's $< .001$, $M_r = .59$) were all moderately to strongly correlated. Thus, the composite of the ratings provided by the three research assistants for each rating among

each participant were averaged together for use in primary analyses. Further correlational analyses indicated that research assistant perceptions of the degree to which participants admitted bias were highly correlated with their own subjective perceptions of bias ($r = .82, p < .001$). Thus, the mean of these ratings was used to represent the degree to which participants admitted bias within their introspective responses in primary analyses (henceforth referred to as Introspected Prejudice).

Preliminary Analyses

Regarding overall trait ratings aggregated across race, a one-way ANOVA indicated that the five experimental conditions did not differ with regard to overall mean trait ratings toward critical targets, $F(4, 210) = 1.67, p = .16$. Additionally, no statistically significant differences emerged between experimental conditions regarding trait ratings toward overall individual targets, $F(4, 210) = 1.69, p = .15$. A marginally significant difference emerged between experimental conditions regarding overall trait ratings toward group targets, $F(4, 210) = 2.17, p = .08$. Tukey post-hoc tests revealed that participants in the Control condition demonstrated marginally lower overall group target trait ratings ($M = 4.89, SD = .54$) than those in the Self-Affirmation Only condition ($M = 5.23, SD = .65, p = .09$). No differences emerged in overall ratings toward Black targets, $F(4, 210) = 1.44, p = .22$, or White targets, $F(4, 210) = 1.99, p = .10$, as a function of experimental condition.

No differences between experimental conditions in pro-Black individual trait ratings, $F(4, 210) = .19, p = .94$, or group trait ratings emerged, $F(4, 210) = .99, p = .42$. However, a marginally significant difference emerged between experimental conditions

regarding pro-Black overall trait ratings, $F(4, 210) = 2.25, p = .07$. Tukey post-hoc tests revealed that participants in the Introspection Only condition ($M = .13, SD = .25$) expressed lower pro-Black overall trait ratings than those in the Introspection First/Self-Affirmation Second condition ($M = .32, SD = .32, p = .05$). Automatic attitude estimates and motivation factor scores did not vary by condition.

Consistent with previous research (Zabel & Olson, 2014), Concern motivation was correlated with overall trait ratings toward group ($r = .34, p < .001$) and individual ($r = .30, p < .001$) targets. In comparison, Restraint motivation was not correlated with overall trait ratings toward individual ($r = -.01, p = .94$) or group ($r = -.00, p = .96$) targets. Consistent with previous research (Zabel & Olson, 2014), Concern was positively correlated with overall trait ratings ($r = .58, p < .001$), as well as overall trait ratings toward individual ($r = .56, p < .001$) and group ($r = .56, p < .001$) targets among participants in the Control condition. Furthermore, Concern was positively correlated with overall trait ratings toward Black ($r = .61, p < .001$) and White ($r = .48, p < .001$) targets, as well as ratings toward Black individual ($r = .57, p < .001$), Black group ($r = .59, p < .001$), White individual ($r = .47, p < .001$), and White group ($r = .45, p < .001$) targets among Control condition participants. Consistent with previous research (Zabel & Olson, 2014), Concern was weakly positively correlated with pro-Black group trait ratings ($r = .22, p = .16$) among Control condition participants. Furthermore, Concern was positively correlated with pro-Black individual trait ratings ($r = .24, p = .12$), as well as pro-Black overall trait ratings ($r = .26, p < .10$), albeit not significantly. These correlations are generally congruent with the “like everyone” strategy that high Concern

individuals tend to employ (Zabel & Olson, 2014). In comparison, in the Control condition, Restraint was not significantly correlated with any of the trait ratings, be they group, individual, Black, White, or a difference score comparing races (all r 's $< |.17|$, p 's $> .29$).

Primary Analyses

Analysis strategy. First, I report on the degree to which participants' introspected prejudices were associated with automatic racial attitudes across conditions in which introspection occurred. These analyses will reveal the degree to which individuals were aware of and willing to report their automatic racial biases across experimental conditions. I then conduct analyses to examine the degree to which automatically-activated racial attitudes, Concern, and experimental conditions interact to predict overall pro-Black trait ratings, followed by looking at pro-Black trait ratings toward groups and individuals separately. I next report on how automatically-activated racial attitudes, Concern, and experimental conditions interact to predict self-reported prejudice (i.e., the composite of subtle prejudice, blatant prejudice, anti-Black attitudes, and symbolic racism). I then examine the interaction of automatically-activated racial attitudes, Restraint, and experimental conditions in predicting overall pro-Black trait ratings, pro-Black individual and group trait ratings separately, and self-reported prejudice.

Admission of racial biases. One primary question of the current work is whether individuals have introspective access as to their racial biases and the conditions under which they are willing to report them. Indeed, high Concern individuals fail to correct for their prejudices in previous research (Zabel & Olson, 2014), which is consistent with a

lack of introspective access to their own prejudices, and/or a willingness to report them. Thus, I first examined the degree to which participant automatic attitude estimates were correlated with research assistant perceptions of participant admittance of bias (i.e., Introspected Prejudice), as indicated by coding of participant introspection responses.

I conducted a hierarchical regression utilizing dummy coded predictors with the Self-Affirmation First/Introspection Second condition coded as the reference group and the Introspection First/Self-Affirmation Second and Introspection Only conditions coded as predictor variables. Automatically-Activated Racial Attitudes and Concern were also entered as predictor variables and Introspected Prejudice was entered as the dependent variable. The proper two- and three-way interactions of these terms were entered as well. The regression uncovered that Automatically-activated racial attitudes predicted Introspected Prejudice in the Self-Affirmation First/Introspection Second condition ($r = .26, p < .09$), but not the Introspection First/Self-Affirmation Second condition ($r = -.16, p = .34$), $t(115) = -1.97, p = .05$, or Introspection Only condition ($r = -.02, p = .90$), $t(121) = -1.15, p = .25$. Additionally, the difference in the correlation between automatically-activated attitudes and introspected prejudice among those in the Self-Affirmation First/Introspection Second condition (relative to Introspection First/Self-Affirmation Second or Introspection Only conditions) did not vary as a function of Concern (p 's $> .55$). These findings suggest that individuals do have some degree of accurate introspective access as to their own prejudices and are willing to admit them, at least, as hypothesized, in contexts in which self-threat may be reduced by a self-affirmation.

The aforementioned analyses provide some initial evidence that individuals may have access as to their own prejudices and a willingness to express them in contexts when self-threat is first reduced by a self-affirmation. Interestingly, the Self-Affirmation First/Introspection Second condition uniquely allowed for accurate introspective access relative to the Introspection First/Self-Affirmation Second condition. Thus, the Introspection First/Self-Affirmation Second and Self-Affirmation First/Introspection Second conditions were treated as separate conditions in all subsequent experimental analyses (as opposed to collapsing across conditions). Based on these initial findings that Self-Affirmation First/Introspection Second condition participants have more accurate access as to their automatic racial attitudes relative to other experimental conditions, I next examined whether individuals in the Self-Affirmation First/Introspection Second condition were likely to correct for their prejudices in their reported impressions of Black and White targets, especially as hypothesized when Concern and automatically-activated racial prejudices were high.

Prejudice correction on trait ratings. Dummy coding was utilized to compare each of the five experimental conditions with regard to the hypothesized Concern x Automatically-Activated Racial Attitudes interaction in predicting pro-Black trait ratings to determine whether self-affirmation, introspection, or a combination of both elements may be most efficacious at leading to prejudice correction among egalitarian-oriented individuals. To test primary hypotheses, hierarchical regressions with overall pro-Black trait ratings entered as the dependent variable were conducted. In the first regression, the Self-Affirmation First/Introspection Second condition was coded as the reference group.

In the second, third, and fourth regressions, the Control, Introspection Only, and Self-Affirmation Only conditions were coded as the reference groups, respectively. In each regression, the four predictor variables, as well as Concern and automatically-activated racial attitudes were entered in Step 1. Concern and automatically-activated attitudes were centered prior to forming the product terms of the interaction entered in this and all subsequently reported regressions to reduce multicollinearity among the predictors (Aiken & West, 1991). In Step 2, the two-way interactions of Concern with each of the four predictor variables, as well as automatically-activated racial attitudes with each of the four predictor variables were entered within each regression. The Automatically-Activated Racial Attitudes x Concern interaction was also entered in Step 2 within each regression. In Step 3, the three-way interactions of each of the four predictor variables with Concern and automatically-activated racial attitudes were entered within each regression.

The regressions revealed significant Predictor main effects (t 's $> |2.13|$, p 's $< .04$), such that those in the Introspection Only condition ($M = .13$, $SD = .25$) expressed lower overall pro-Black trait ratings than participants in in the Self-Affirmation First/Introspection Second condition ($M = .27$, $SD = .26$), Introspection First/Self-Affirmation Second condition, ($M = .32$, $SD = .32$), Self-Affirmation Only condition ($M = .26$, $SD = .28$), or Control condition ($M = .26$, $SD = .39$). No other main effects emerged (all p 's $> .35$). However, a Predictor x Concern interaction emerged, $t(195) = -2.13$, $p = .03$. The nature of the interaction was such that among those in the Control condition, Concern was positively, but not significantly correlated with overall pro-Black

trait judgments, $r(42) = .26, p < .10$. This finding is congruent with previous research (Zabel & Olson, 2014). However, among those in the Introspection Only condition, Concern was negatively, but not significantly correlated with overall pro-Black trait ratings, $r(43) = -.20, p = .20$. No other two-way interactions emerged significantly (all p 's $> .13$).

However, several three-way interactions emerged significantly. Specifically, as expected, the Predictor (Control vs. Self-Affirmation First/Introspection Second) x Automatically-Activated Racial Attitudes x Concern, $t(191) = -2.12, p = .04$, as well as the Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation First/Introspection Second) x Automatically-Activated Racial Attitudes x Concern interaction, $t(191) = -2.29, p = .02$, emerged significantly. Furthermore, a marginally significant Predictor (Self-Affirmation Only vs. Self-Affirmation First/Introspection Second) x Automatically-Activated Racial Attitudes x Concern interaction emerged, $t(191) = -1.66, p < .10$. No other three-way interactions emerged (all p 's $> .19$).

The nature of these interactions was such that the Automatically-Activated Racial Attitudes x Concern interaction emerged significantly in the Self-Affirmation First/Introspection Second condition (see Figure 1), $t(41) = 2.27, p = .03$, but not in the Control (see Figure 2), $t(38) = -.87, p = .39$, Introspection First/Self-Affirmation Second (see Figure 3), $t(33) = -1.60, p = .12$, or Self-Affirmation Only condition, $t(40) = -.48, p = .64$. Simple slope analyses indicated that among participants in the Self-Affirmation First/Introspection Second condition with more pro-White automatically-activated attitudes (+1 SD), Concern was expectedly positively associated with overall pro-Black

trait ratings, $t(41) = 2.25, p = .03$. Among those with more pro-Black automatically-activated attitudes (-1 SD), Concern was not correlated with overall pro-Black trait ratings, $t(41) = -.79, p = .43$. Thus, as expected, as pro-White automatic attitudes increased, Concern motivation led to corrections for automatic-prejudice among those who were self-affirmed first and then introspected on their racial biases. Moreover, this pattern of correction for automatic prejudice among high Concern participants was different from experimental conditions in which no self-affirmation or introspection occurred, or when self-affirmation followed racial bias introspection. This is congruent with my reasoning that self-affirmation may reduce the self-threat to introspecting about one's own racial biases, which may in turn facilitate deeper and more accurate racial attitude introspection and prejudice correction effects among high Concern individuals.

Results thus far indicate that self-affirmation and racial bias introspection are necessary to lead to corrections for automatic prejudice in the form of overall pro-Black trait ratings toward among high egalitarian-oriented (Concern) individuals. However, previous research suggests that given its underlying tenets, Concern may be especially likely to be activated and hence influential in prejudice correction in social judgments toward *group* (relative to individual) targets (Olson & Fazio, 2009). Thus, I next examined whether the expected pattern of automatic prejudice correction among high Concern individuals previously uncovered in the Self-Affirmation First/Introspection Second condition was more apparent toward group, relative to individual targets.

To do so, I conducted a series of four hierarchical regressions in the same manner as that of the previously reported regressions. The only difference was that pro-Black

group trait ratings were entered as the dependent variable. The regressions yielded a main effect, $t(183) = 2.04, p = .04$, such that participants in the Introspection Only condition ($M = .32, SD = .35$) demonstrated lower pro-Black group trait ratings than those in the Introspection First/Self-Affirmation Second condition ($M = .51, SD = .33$). No other main effects emerged significantly across regressions (all p 's $> .12$). Moreover, no two-way interactions between study variables emerged (all p 's $> .10$). However, the previously elaborated three-way interactions emerged and mirrored results from the regression with overall pro-Black trait ratings entered as the dependent variable. Specifically, the Predictor (Control vs. Self-Affirmation First/Introspection Second) x Automatically-Activated Racial Attitudes x Concern interaction emerged significantly, $t(170) = -2.07, p = .04$, as did the Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation First/Introspection Second) x Automatically-Activated Racial Attitudes x Concern interaction, $t(170) = -2.25, p = .03$.

The nature of these interactions mirrored the previously elaborated results, with the Concern x Automatically-Activated Racial Attitudes interaction emerging in the Self-Affirmation First/Introspection Second condition, $t(41) = 2.19, p = .03$. Simple slope analyses indicated that among participants in the Self-Affirmation First/Introspection Second condition with more pro-White automatically-activated attitudes (+1 SD), Concern was marginally positively associated with pro-Black group trait ratings, $t(41) = 1.90, p = .06$. Among those with more pro-Black automatically-activated attitudes (-1 SD) in this condition, Concern was not associated with pro-Black group trait ratings, $t(41) = -1.02, p = .31$. Thus, as expected, as automatically-prejudiced attitudes increased,

Concern led to corrections for automatic-prejudice in the form of pro-Black group trait ratings among those who were self-affirmed first and then introspected on their racial biases. The Concern x Automatically-Activated Racial Attitudes interaction also emerged marginally significantly in the Introspection First/Self-Affirmation Second condition, $t(33) = -1.78, p = .08$. Simple slope analyses indicated that among participants with pro-White (+1 SD) automatically-activated racial attitudes in the Introspection First/Self-Affirmation Second condition, Concern negatively, albeit not significantly, was correlated with pro-Black group trait ratings, $t(33) = -1.30, p = .20$. Among participants with more pro-Black (-1 SD) automatically-activated racial attitudes (-1 SD) in this condition, Concern was not associated with pro-Black group trait ratings, $t(33) = 1.45, p = .16$.

I next examined whether this pattern of prejudice correction among high Concern participants was also present toward individual targets by conducting a series of four hierarchical regressions using the same hierarchical regression approach as previously elaborated. In each regression, pro-Black individual trait ratings were entered as the dependent variable. These regressions revealed no statistically significant main effects (all p 's > .15) or two-way interactions (all p 's > .13). Interestingly, no three-way interactions emerged significantly (all p 's > .11), although patterns were similar in nature to results reported with pro-Black overall trait ratings entered as the dependent variable. Thus, the pattern of prejudice correction among high Concern participants in the Self-Affirmation First/Introspection Second condition relative to other experimental conditions seems to be especially salient in group target construal contexts.

These results clearly indicate that a self-affirmation, followed by introspection of racial biases, leads to social judgment corrections for automatic prejudice among egalitarian-oriented Whites. Moreover, order of these tasks matters; racial introspection followed by a self-affirmation is not sufficient to producing corrections for automatic prejudices among high Concern individuals, and may even lead to more pro-White judgments or behaviors. Furthermore, congruent with previous research (Olson & Fazio, 2009), Concern appears to be especially applicable to prejudice correction in group target construal contexts. I next examined potential converging evidence for prejudice correction among high Concern individuals in the Self-Affirmation First/Introspection Second condition regarding self-reported measures of prejudice.

Prejudice correction regarding explicit attitude judgments. Self-report measures of subtle prejudice, blatant prejudice, anti-Black attitudes, and symbolic racism were positively correlated (all r 's $> .42$, all p 's $< .001$, $M_r = .58$). As such, z scores of these measures were computed. The mean of these z scores was used as a composite measure of self-reported prejudice in subsequent primary analyses. These composite self-reported prejudice scores ranged from -2.05 to 2.18. To further explore the nature of prejudice correction effects, regressions with the prejudice composite measure entered as the dependent variable were conducted. Specifically, I conducted a series of four hierarchical regressions in a similar manner to that of the previously reported regressions. In each regression, prejudice composite scores were entered as the dependent variable.

Regressions indicated a main effect of Concern, $t(204) = -5.77$, $p < .001$, such that Concern was negatively predictive of the self-reported prejudice composite. No other

main effects emerged (all p 's > .15). However, the Predictor (Introspection First/Self-Affirmation Second vs. Control) x Concern interaction, $t(195) = 1.83, p = .07$, Predictor (Introspection First/Self-Affirmation Second vs. Introspection Only) interaction, $t(195) = 1.95, p = .05$, and Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation Only) x Concern interaction, $t(195) = 1.90, p = .06$, each emerged. The nature of these interactions was that Concern was negatively associated with self-reported prejudice among Introspection Only ($r = -.47, p < .01$), Control ($r = -.46, p < .01$), and Self-Affirmation Only ($r = -.46, p < .01$) condition participants, whereas it was not among Introspection First/Self-Affirmation Second condition participants ($r = -.07, p = .69$).

Additionally, the Predictor (Introspection First/Self-Affirmation Second vs. Introspection Only) x Automatically-Activated Racial Attitudes, $t(185) = 1.83, p = .07$, as well as Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation Only) x Automatically-Activated Racial Attitudes interaction, $t(195) = 1.70, p = .09$, each emerged marginally significantly. The nature of these interactions was that automatically-activated racial attitudes were positively correlated with self-reported prejudice among Introspection First/Self-Affirmation Second condition participants ($r = .31, p = .06$), but not among participants in the Introspection Only ($r = -.11, p = .48$) or Self-Affirmation Only condition ($r = -.02, p = .88$). No other two-way interactions emerged significantly (all p 's > .11).

The Predictor (Self-Affirmation First/Introspection Second vs. Introspection First/Self-Affirmation Second) x Concern x Automatically-Activated Racial Attitudes, t

(191) = 1.88, $p = .06$, as well as the Predictor (Self-Affirmation First/Introspection Second vs. Self-Affirmation Only) x Concern x Automatically-Activated Racial Attitudes interaction, $t(191) = 1.82$, $p = .07$, each emerged marginally significantly. No other three-way interactions emerged significantly (all p 's > .11). The Concern x Automatically-Activated Racial Attitudes interaction trended toward significance in predicting self-reported prejudice in the Self-Affirmation First/Introspection Second condition, $t(41) = -1.56$, $p = .13$. Simple slope analyses indicated that among individuals with pro-White automatically-activated attitudes (+1 SD) in the Self-Affirmation First/Introspection Second condition, Concern was as expected negatively predictive of self-reported prejudice composite scores, $t(41) = -2.42$, $p = .02$. In comparison, among individuals with pro-Black (-1 SD) automatically-activated attitudes in this condition, Concern was not predictive of self-reported prejudice, $t(41) = -.28$, $p = .78$.

The Concern x Automatically-Activated Racial Attitudes interaction trended toward significance in predicting self-reported prejudice in the Introspection First/Self-Affirmation Second, $t(33) = 1.49$, $p = .15$. Simple slope analyses also indicated that among individuals with pro-White automatically-activated attitudes (+1 SD) in the Introspection First/Self-Affirmation Second condition, Concern was positively associated with self-reported prejudice, albeit not significantly, $t(33) = .77$, $p = .45$. In comparison, among individuals with pro-Black automatically-activated attitudes (-1 SD) in this condition, Concern was negatively, albeit not significantly associated with self-reported prejudice, $t(33) = -1.41$, $p = .17$. The Concern x Automatically-Activated Racial

Attitudes interaction did not emerge significantly in the Self-Affirmation Only condition, $t(40) = .84, p = .41$.

These findings indicate that self-affirmation followed by racial attitude introspection appears to prompt individuals with pro-White automatically-activated racial attitudes to reduce self-reported prejudice as Concern motivation increases. These results are congruent with the direct test of primary hypotheses reported earlier, in which individuals with pro-White automatically-activated attitudes corrected for their automatic prejudices in the form of pro-Black trait ratings as Concern Motivation increased in the Self-Affirmation First/Introspection Second condition. Results thus far indicate that self-affirmation must precede racial prejudice introspection in order for prejudice correction among high Concern individuals to be facilitated. As a final note, prejudice correction effects among high Concern individuals in the Self-Affirmation First/Introspection Second condition do not appear to be strictly limited toward individual or group targets in the social perception process, but rather are applicable to social perception, judgments, and attitudes more broadly.

Role of restraint motivation in trait judgment prejudice correction. Given that Concern motivation is more intrinsically-based and tied to self-image than Restraint motivation, I expected Restraint to similarly facilitate prejudice correction across all experimental conditions, including the Control condition. Indeed, the Control condition provides an opportunity for a conceptual replication of my previous research (Zabel & Olson, 2014) in which high Restraint individuals correct for their prejudices, especially in individual target construal situations. However, an examination of the Restraint x

Automatically-Activated Racial Attitudes interaction among participants in the Control condition of the current study with overall pro-Black ratings entered as the dependent variable indicated no statistically significant main effects (all p 's $> .22$) or the expected interaction, $t(38) = -.15, p = .88$. Furthermore, the Restraint x Automatically-Activated Racial Attitudes interaction did not emerge regarding pro-Black individual ($p = .75$) or group ($p = .95$) trait ratings. While at first blush these results indicate a failure to replicate, the Control condition in the current study was uniquely different from my previous work in that in the current work, participants were informed that racial biases are pervasive across society prior to providing trait ratings to maintain consistency across all conditions. Perhaps this prompt diminished activation of their conflict avoidance motive (i.e., Restraint), either by overactivating Concern motivation or providing external justification that reduced the activation of Restraint.

Speculation aside, I next sat out to verify that the interaction of Automatically-Activated Racial Attitudes and motivation (i.e., Concern) to control prejudiced reactions among individuals in the Self-Affirmation First/Introspection Second condition was a uniquely Concern (and not Restraint) facilitated effect. The same hierarchical approach using four regressions employed in previous analyses was utilized regarding the variables of Restraint, automatically-activated racial attitudes, and predictor variables. In each regression, overall pro-Black trait ratings were entered as the dependent variable. No Restraint main effect, $t(204) = -.09, p = .93$, or Restraint x Automatically-Activated Racial Attitude, $t(195) = -1.22, p = .22$, emerged. However, the Predictor (Introspection First/Self-Affirmation Second vs. Control) x Restraint interaction emerged, $t(195) =$

2.47, $p = .01$. The nature of the interaction was such that among those in the Introspection First/Self-Affirmation Second condition, Restraint was marginally positively associated with overall pro-Black trait ratings ($r = .32$, $p = .06$), whereas this correlation was negative but not statistically significant among Control condition participants ($r = -.17$, $p = .29$). No other two-way (all p 's $> .14$) or three-way (all p 's $> .50$) interactions emerged significantly.

Results indicate that Restraint does not interact with automatically-activated racial attitudes in predicting overall pro-Black trait ratings. More importantly regarding the current study's purposes, results thus far indicate that as expected, Restraint does not differentially facilitate prejudice correction effects between any of the experimental conditions. However, my previous research (Zabel & Olson, 2014) indicates that Restraint is activated more strongly and hence more applicable to prejudice correction in individual target construal contexts. Thus, I next examined the degree to which Restraint interacted with automatically-activated racial attitudes to predict pro-Black individual target ratings across and between experimental conditions. To do so, and as previously elaborated, I conducted four hierarchical regressions with pro-Black individual trait ratings entered as the dependent variable.

No Restraint main effect, $t(204) = .93$, $p = .35$, or Restraint x Automatically-Activated Racial Attitudes interaction emerged, $t(195) = -1.00$, $p = .32$. However, the Predictor (Introspection First/Self-Affirmation Second vs. Control) x Restraint interaction emerged, $t(195) = 3.31$, $p = .01$, as did the Predictor (Introspection First/Self-Affirmation Second vs. Introspection Only) x Restraint interaction (albeit marginally), $t(195) = 1.82$,

$p = .07$, the Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation Only) x Restraint interaction (albeit marginally), $t(195) = 1.76$, $p = .08$, and the Restraint x Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation First/Introspection Second) interaction, $t(195) = 2.00$, $p < .05$. The nature of these interactions was such that Restraint was positively associated with pro-Black individual trait ratings among Introspection First/Self-Affirmation Second condition participants ($r = .44$, $p < .01$), but not among participants in the Control condition ($r = -.16$, $p = .30$), Introspection Only condition ($r = .10$, $p = .52$), Self-Affirmation Only condition ($r = .10$, $p = .52$), or Self-Affirmation First/Introspection Second condition ($r = .01$, $p = .92$). No other previously unreported two-way (all p 's $> .12$) or three-way (all p 's $> .75$) interactions emerged significantly.

I also examined how Restraint, automatically-activated racial attitudes, and experimental conditions interacted to predict pro-Black group trait ratings using the previously elaborated hierarchical regression strategy. These regressions yielded no Restraint main effects, $t(183) = -.78$, $p = .44$, or Restraint x Automatically-Activated Racial Attitudes interaction, $t(174) = -.78$, $p = .44$. Furthermore, no two-way (all p 's $> .25$) or three-way (all p 's $> .46$) interactions emerged significantly. Together, these regression reveal no differences in how Restraint interacts with automatically-activated racial attitudes to predict pro-Black trait ratings between experimental conditions. Thus, the efficacy of self-affirmation followed by racial prejudice introspection in leading to prejudice correction appears to be specific to high Concern (and not Restraint) individuals. This pattern of findings is congruent with my theoretical reasoning regarding

the efficacy of self-affirmation in reducing the threat of racial attitude introspection to the self among egalitarian-oriented individuals, as well as the underlying tenets of Concern and Restraint motives.

Role of restraint motivation in self-reported racial attitude prejudice

correction. To further inspect the potential nature of Restraint in prejudice correction effects, I initially examined the role of Restraint in leading to self-reported prejudice correction effects among individuals in the Control condition. However, an examination of the Restraint x Automatically-Activated Racial Attitudes interaction among participants in the Control condition with the self-reported prejudice composite entered as the dependent variable indicated no statistically significant main effects (all p 's $> .80$) or the expected interaction, $t(38) = -.09, p = .93$. Thus, as previously elaborated, these results indicate a failure to replicate previous research (Zabel & Olson, 2014), although I suspect this may be due to slight differences in the methodology between the current study and this past research.

I next conducted four separate regressions, each with the self-reported composite prejudice measure entered as the dependent variable. In each regression, the predictor variables, Restraint, and automatically-activated racial attitudes, as well as the proper two-way and three-way interactions of these terms were entered. Given Restraint's underlying tenets, I did not expect it to interact with automatic racial attitudes differentially between experimental conditions. Results indicated that neither the Restraint main effect, $t(204) = -.12, p = .91$, or Restraint x Automatically-Activated Racial Attitudes interaction, $t(195) = -.26, p = .80$, emerged significantly. However, the

Predictor (Self-Affirmation First/Introspection Second vs. Introspection Only) x Automatically-Activated Racial Attitudes interaction emerged significantly, $t(195) = -1.94, p = .05$. For participants in the Self-Affirmation First/Introspection Second condition, as automatically-activated racial attitudes became more pro-White, self-reported prejudice levels also tended to increase ($r = .28, p = .06$). This finding is congruent with previously-reported findings that individuals may have greater access to their automatic prejudices under conditions of self-affirmation followed by racial attitudes introspection. In comparison, among Introspection Only condition participants, as automatically-activated racial attitudes became pro-White, self-reported prejudice levels tended to decrease, albeit not significantly ($r = -.11, p = .48$). No other two-way (all p 's $> .12$) or three-way (all p 's $> .45$) interactions emerged significantly. Thus, correction of automatic prejudice among highly motivated individuals in the Self-Affirmation First/Introspection Second condition relative to the Control condition appears to be, as expected, a Concern-fueled effect.

Supplementary Analyses

Trait rating response latencies. If participants with pro-White automatically-activated attitudes took longer to make trait ratings toward critical targets as Concern motivation increased in the Self-Affirmation First/Introspection Second condition but not the Control (Control) condition, it would provide converging evidence for more deliberative processes (Olson & Fazio, 2009) important to automatic prejudice correction. To assess this possibility, I conducted a hierarchical regression in which the mean of critical overall trait rating trial response latencies was entered as the dependent variable.

The Self-Affirmation First/Introspection Second condition was dummy coded as the reference group, with corresponding predictor variables entered, as well as the necessary two- and three-way interactions of variables.

The regression yielded a marginally significant Predictor (Introspection First/Introspection Second) main effect, $t(204) = 1.92, p = .06$, such that those in the Introspection First/Self-Affirmation Second condition ($M = 2,148.33, SD = 484.12$) took longer than those in the Self-Affirmation First/Introspection Second condition ($M = 1,948.37, SD = 512.24$). No other statistically significant main effects (all p 's $> .12$) or two-way interactions (all p 's $> .19$) emerged. However, the expected three-way Predictor (Control vs. Self-Affirmation First/Introspection Second) x Concern x Automatically-Activated Racial Attitudes interaction emerged, $t(191) = -3.90, p < .001$.

The nature of this interaction was such that among those in the Self-Affirmation First/Introspection Second condition, the Concern x Automatically-Activated Racial Attitudes interaction emerged significantly (see Figure 4), $t(41) = 2.57, p = .01$. Simple slope analyses indicated that among those in the Self-Affirmation First/Introspection Second condition with pro-White automatically-activated racial attitudes (+1 SD), Concern was expectedly positively associated with overall critical trial trait rating response latencies, albeit not significantly $t(41) = 1.35, p = .19$. However, among those with pro-Black automatically-activated racial attitudes (-1 SD) in this condition, Concern was negatively predictive of overall critical trial trait rating response latencies, $t(41) = -2.03, p < .05$. In the Control condition, the Concern x Automatically-Activated Racial Attitudes interaction also emerged significantly, but with a different pattern (see Figure

5), $t(38) = -2.92, p < .01$. Simple slope analyses indicated that among participants in the Control condition with pro-White automatically-activated racial attitudes (+1 SD), Concern was negatively predictive of overall critical trial trait rating response latencies, $t(38) = -3.39, p < .01$. However, among participants with pro-Black automatically-activated racial attitudes (+1 SD) in the Control condition, Concern was not predictive of overall critical trial trait rating response latencies, $t(38) = -.43, p = .67$. The Predictor (Self-Affirmation Only vs. Self-Affirmation First/Introspection Second) x Concern x Automatically-Activated Racial Attitudes interaction also emerged significantly, $t(191) = -2.52, p = .01$. The nature of this interaction was elaborated above regarding Self-Affirmation First/Introspection Second condition participants. The Concern x Automatically-Activated Racial Attitudes interaction did not emerge significantly among Self-Affirmation Only condition participants, $t(40) = -1.12, p = .27$ (see Figure 6).

The finding that high Concern individuals with pro-White automatic attitudes in the Control condition took significantly less time to make trait judgments is congruent with the previously elaborated “like everyone” heuristic and lack of prejudice correction among high Concern individuals with pro-White automatic attitudes demonstrated in previous research (Zabel & Olson, 2014). Furthermore, the finding that high Concern individuals with pro-White automatic attitudes in the Self-Affirmation First/Introspection Second condition took *more* time to make trait judgments is congruent with my reasoning that self-affirmation reduces the self-threat of introspection, which in turn leads to reasonably accurate introspection and anticipated prejudice correction effects via more deliberative processing. Indeed, Control condition participants high in Concern may fail

to introspect on their racial prejudices to avoid a self-threat, which in turn leads them to employ the “like everyone” heuristic congruent with their egalitarian focus but not to correct for their automatic racial prejudices. The aforementioned response latency findings are congruent with this reasoning.

I next conducted two regressions to examine whether the Concern x Automatically-Activated Racial Attitudes interactions in the Self-Affirmation First/Introspection Second and Control conditions were driven by response latencies toward critical Black targets, critical White targets, or both. In each regression, response latencies regarding overall trait rating trials toward all critical Black targets and all critical White targets were entered as the dependent variable, respectively. In both regressions, the Self-Affirmation First/Introspection Second condition was coded as the reference group. Variables were entered in the regressions in the same manner as elaborated on previously.

The first yielded the same previously reported Predictor (Self-Affirmation First/Introspection Second vs. Control) x Concern x Automatically-Activated Racial Attitudes interaction, $t(191) = -3.63, p < .001$, and Predictor (Self-Affirmation First/Introspection Second vs. Self-Affirmation Only) x Concern x Automatically-Activated Racial Attitudes interaction, $t(191) = -2.59, p = .01$. The nature of these interactions mirrored that with the overall response latencies toward critical targets entered as the dependent variable. Specifically, among Self-Affirmation First/Introspection Second participants with pro-White (+1 SD) automatically-activated racial attitudes, Concern was positively associated with response latencies toward Black

targets, whereas among those in the Control condition, Concern was negatively associated with response latencies toward Black targets. These findings suggest that the same patterns elaborated on earlier regarding the “like everyone” heuristic in the Control (control) condition and more deliberative prejudice correction in the Self-Affirmation First/Introspection Second condition among high Concern individuals are supported by participants’ response latencies toward Black critical targets. Further follow-up analyses indicated that the Predictor x (Self-Affirmation First/Introspection Second vs. Control) x Concern x Automatically-Activated Racial Attitudes interaction emerged both toward Black group and individual targets (p ’s < .01).

The second regression yielded the same Predictor (Self-Affirmation First/Introspection Second vs. Control) x Concern x Automatically-Activated Racial Attitudes interaction reported previously, $t(191) = -4.02, p < .001$, as well as the Predictor (Self-Affirmation First/Introspection Second vs. Self-Affirmation Only) x Concern x Automatically-Activated Racial Attitudes interaction, $t(191) = -2.37, p = .02$. The nature of these interactions mirrored those elaborated on earlier. Specifically, among Self-Affirmation First/Introspection Second condition participants with pro-White (+1 SD) automatically-activated racial attitudes, Concern was positively associated with response latencies toward White targets, albeit not significantly, $t(41) = 1.46, p = .15$, whereas among Control condition with pro-White (+1 SD) automatically-activated racial attitudes, Concern was negatively predictive of response latencies toward White targets, $t(38) = -3.31, p < .01$. Follow-up analyses indicated that the Predictor x (Self-Affirmation First/Introspection Second vs. Control) x Concern Motivation x Automatically-Activated

Racial Attitudes interaction emerged both toward White group and individual targets (p 's $< .01$).

These findings provide response latency support for the “like everyone” heuristic and failure to correct for automatic prejudices among high Concern individuals in previous research (Zabel & Olson, 2014) and suggest that this heuristic influences judgments both toward Black and White targets among high Concern individuals. Furthermore, the response latencies provide support for more deliberative social judgment in the Self-Affirmation First/Introspection Second condition, which converges with the prejudice correction trait rating effects demonstrated among individuals with pro-White automatically-activated racial attitudes in the Self-Affirmation First/Introspection Second condition as Concern Motivation increases.

Chapter 5: Discussion

Automatic prejudices are pervasive. Thus, it is important to understand the processes by which automatic prejudices can be corrected for in social judgments and behaviors, which in turn may hold direct implications for improving intergroup relations and outcomes for lower status group members. Previous research (Fazio, 1990; Fiske et al., 1999) indicates that motivation and opportunity (i.e., time, attentional resources) are essential to enabling automatic prejudice correction. However, even when both motivation and opportunity exist, automatic prejudice correction may not occur (Zabel & Olson, 2014). This previous research indicates that high egalitarian-oriented (Concern) Whites fail to correct for their automatically-activated racial attitudes in social perception.

Two possibilities for this tendency were raised and tested in this work. The first possibility was that egalitarian-oriented individuals may lack introspective access to their automatic racial attitudes (Greenwald & Banaji, 1995; but see Fazio & Olson, 2003), rendering automatic prejudice correction impossible. However, previous research (Phillips & Olson, 2014), as well as the finding in the current study that automatically-activated racial attitudes were positively correlated with the degree to which participants admitted bias in the Self-Affirmation First/Introspection Second condition contradicts this possibility. Thus, clearly individuals do have the capacity to accurately introspect on their automatic racial attitudes under the right conditions (i.e., self-affirmation followed by racial attitude introspection).

This finding suggests that egalitarian-oriented individuals in previous work (e.g., Zabel & Olson, 2014) in which no self-affirmation or racial attitude introspection occurred may have failed to correct for their automatic prejudices due to lacking awareness of their automatic racial attitudes, but not the *ability* to introspect on their racial attitudes. The question then becomes: Why do egalitarian-oriented individuals fail to reflect on and subsequently correct for their automatic prejudices? Egalitarian-oriented (high Concern) individuals have a non-prejudiced self-image. Thus, the prospect of introspecting on their racial prejudices may be a threat to their self-image, a threat that previous research would suggest that egalitarian-oriented individuals are motivated to avoid (Gaertner & Dovidio, 1986). Consistent with a motivational perspective and prominent theoretical approaches (e.g., Aversive Racism Theory; Gaertner & Dovidio, 1986), egalitarian-oriented Whites may fail to introspect and subsequently correct for their automatic prejudices in judgments in behavior under normal circumstances in order to avoid a threat to their egalitarian self-image. Given its success in facilitating attention to otherwise self-threatening information, I proposed that self-affirmation (Steele, 1988) would enhance the likelihood that egalitarian-oriented Whites reflected accurately on their automatic racial biases. Furthermore, I expected that self-affirmation followed by racial attitude introspection would prompt automatic prejudice correction among egalitarian-oriented Whites in a manner congruent with prominent bias correction theories (Wegener & Petty, 1997) and social psychological research findings (Olson & Fazio, 2004; 2014).

Congruent with this reasoning, high Concern individuals in the Self-Affirmation First/Introspection Second condition corrected for their automatic prejudices in the form of trait judgments, whereas those in the Control condition did not. These findings are substantiated by response latency data indicating that individuals with pro-White automatic attitudes took longer in trait rating judgments toward critical targets as Concern increased when they self-affirmed and then introspected. These findings add convergent validity for automatic prejudice correction in the Self-Affirmation First/Introspection Second condition by indicating that judgments made in this context were made more deliberately as individuals corrected for their prejudices. As elaborated earlier, I suspect that self-affirmation reduces the potential self-threat of introspecting about one's own racial biases among egalitarian-oriented Whites, leading to more accurate and deep introspection of one's biases, and subsequently, correction for automatic prejudice.

In the absence of the specific combination of self-affirmation followed by racial attitude introspection, the default strategy among egalitarian-oriented Whites appears to be to "like everyone" using more automatic, heuristic processes (Zabel & Olson, 2014). Replicating previous research, Concern was positively correlated with trait ratings toward all critical targets among participants in the Control condition. As in my previous research (Zabel & Olson, 2014), automatic attitudes did not interact with Concern in predicting pro-Black trait judgments among participants in the Control condition. This finding indicates that the norm is for egalitarian-oriented individuals to fail to account for the influence of their automatically-activated attitudes in social perception, which is congruent with my perspective that egalitarian-oriented individuals fail to correct for

their automatic prejudices due to the self-threat encapsulated in introspecting on such prejudices.

These findings are substantiated by response latency data indicating that high Concern participants made *quicker* trait judgments toward overall critical targets in the Control condition as their automatic attitudes became more pro-White. These reduced response latencies are consistent with the perspective that egalitarian-oriented Whites develop a “like everyone” heuristic to avoid a discrepancy between their automatic prejudice and egalitarian self-image, and that this may be especially true among those with pro-White automatic attitudes whose self-image is most threatened by potentially prejudice-conveying contexts. It is important to note that analyses indicated no interaction of Concern and automatically-activated racial attitudes in predicting overall trait ratings toward all critical targets. Based on the response latency data, one might expect individuals with pro-White automatically-activated attitudes to be especially likely to provide positive trait ratings as Concern increased based on heuristic use. Why this inconsistency exists is unclear. However, I have consistently uncovered no interaction of automatic racial attitudes with Concern in predicting pro-Black and overall trait ratings in this and previous research (Zabel & Olson, 2014). Thus, the anomaly or less substantiated finding that warrants verification and future examination may be that individuals with pro-White attitudes take a shorter amount of time to make judgments as Concern increases.

Interestingly, the automatic prejudice correction pattern among egalitarian-oriented individuals that occurred among Self-Affirmation First/Introspection Second

condition participants did not emerge among participants in the Introspection First/Self-Affirmation Second condition. In contrast, a marginally significant opposite pattern emerged, such that self-affirmation preceded by racial prejudice introspection paradoxically led to *reduced* Pro-Black trait ratings among high Concern individuals with pro-White automatic attitudes. Although unexpected, I suspect this may be due to psychological defensiveness and reactance (Brehm, 1966). Specifically, egalitarian-oriented individuals may be especially likely to respond with disdain and interpret instructions to introspect as accusatory of their own racial prejudices. These individuals may be especially likely to perceive they are being forced to confront their own prejudices outside of their own terms and also when there is no need to do so based on their egalitarian values. Following introspection (which is unlikely to lead to accurate self-reflection based on the data herein), self-affirmations may release participants from negative ruminative thoughts (Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999) regarding being forced to introspect on their racial prejudices. In turn, this may create a backlash effect (Plant et al., 2010) of reactance in which high Concern individuals actually favor White relative to Black targets.

No differences in response latencies among egalitarian-oriented individuals with pro-White automatically-activated attitudes emerged between participants in the Self-Affirmation First/Introspection Second and Introspection First/Self-Affirmation Second conditions. These findings indicate that although the combination of racial prejudice introspection and self-affirmation leads to similar amounts of deliberation in social judgments, the order of these processes is critically important to facilitating automatic

prejudice correction effects. As these results demonstrate, self-affirmation must precede racial prejudice introspection to facilitate prejudice correction effects among high Concern individuals.

The correction for automatic prejudice appears to be especially applicable among high Concern (relative to Restraint) individuals in the Self-Affirmation First/Introspection Second condition toward group targets. This is consistent with previous research indicating that Concern appears to stem from an interest in redressing historical inequalities (Fazio & Olson, 2014), as well as egalitarianism (Fazio & Hilden, 2001), rendering it more applicable and active in impression formation contexts in which targets are construed at the group level (Zabel & Olson, 2014). However, Concern also trended toward facilitating automatic prejudice toward individual targets among those in the Self-Affirmation First/Introspection Second condition. This is also consistent with previous research demonstrating how remarkably flexible construal of social targets can be depending on perceiver motives (Sinclair & Kunda, 1999). In sum, our findings are congruent with previous research indicating that perceiver motives influence construal and social perception processes in the impression context (Kunda & Thagard, 1996), as well as recent research indicating that activation and application of impression-relevant motives (i.e., Concern) is influenced by how a target is construed (Zabel & Olson, 2014).

It is important to note that the prejudice correction effects facilitated by motivation among individuals in the Self Affirmation First/Introspection Second condition appear to be unique to Concern (and not Restraint) motivation. Indeed, this is consistent with Concern being more intrinsically-based and tied to one's self-image than

Restraint, which is more extrinsically-based and interpersonally-activated. Because the self-affirmation manipulation was designed to reduce the self-image threat of racial attitude introspection, I did not expect it to have an influence in producing Restraint-facilitated prejudice correction effects.

Previous research indicates that high Restraint individuals correct for their prejudices when construing targets at the individual level in the absence of racial attitude introspection and self-affirmation (Zabel & Olson, 2014). However, the current study found no interaction of Restraint and automatically-activated attitudes in predicting pro-Black trait judgments across conditions. Although this is a failure to replicate, I expect it may be due to procedural differences between studies. Specifically, in the current research, all participants were informed that automatic prejudices are pervasive in contemporary American society, even if we as individuals are not aware of our own biases. This prompt may have provided high Restraint individuals sufficient extrinsic justification to not correct for their automatic prejudices (i.e., everyone is prejudiced), leading to this failure to replicate previous findings. Another possibility is that the prompt in the current study highlighted race as a social category, facilitated group level target construal, and in turn activated Concern (relative to Restraint) motivation more strongly (Zabel & Olson, 2014), reducing the influence of Restraint in producing prejudice correction effects in the Control condition.

Automatic prejudice correction effects in the Self-Affirmation First/Introspection Second condition are generally consistent with the Moderated Consistency Model (Hofmann et al., 2005), in which awareness teamed with adjustment (motivation) is

posited to reduce implicit-explicit attitude correspondence. In previous examinations (Gschwendner et al., 2006; Hofmann et al., 2005) of the Moderated Consistency Model, awareness has been primarily operationalized as private self-consciousness (Fenigstein et al., 1975). However my experimental focus on racial attitude introspection expands understanding of the role of awareness in at least two ways. First, it examines whether the more conscious action of racial attitude introspection also facilitates adjustment. Furthermore, it provides a more domain specific operationalization of racial attitude awareness, compared to a more generalized, dispositional operationalization of awareness. Indeed, I suspect that generalized self-awareness is distinct from the more conscious and domain specific process of racial attitude introspection. If so, this may provide support for the relative advantage of more nuanced and domain specific operationalizations of awareness in future research examining the roles of awareness and motivation in implicit-explicit consistency.

To assess this possibility in the current research, I examined the degree to which private self-consciousness influenced the extent to which individuals expressed accurate awareness of their automatic racial attitudes. I regressed participants' introspected prejudice on dummy coded predictor variables (Self-Affirmation First/Introspection Second coded as the reference group), automatically-activated racial attitudes, and private self-consciousness, as well as the appropriate two- and three-way interactions of these terms. Importantly, the Private Self-Consciousness x Automatically-Activated Racial Attitudes x Predictor interaction did not emerge, $t(115) = -.06, p = .95$. Furthermore, private self-consciousness was not correlated with either automatic racial attitudes ($r = -$

.10, $p = .51$) or introspected prejudice ($r = -.03$, $p = .85$) in the Self-Affirmation First/Introspection Second condition or any of the other experimental conditions (all r 's $< |.23|$, all p 's $> .14$, $M_r = .12$). Although private self-consciousness has often been used as a dispositional measure of awareness in previous research (Gschwendner et al., 2006; Hofmann et al., 2005), these findings suggest that it plays no role in the degree to which individuals accurately report their own racial biases, and that private self-consciousness may be tapping a more generalized construct distinct from racial attitude awareness specifically. Nevertheless, it is important to reiterate that the current results are congruent with work employing private self-consciousness as an awareness measure, as well as the Moderated Consistency Model (Gschwendner et al., 2006; Hofmann et al., 2005). However, future research should examine the methodological consequences of utilizing more domain specific vs. generalized awareness variables, as well as more active vs. dispositional awareness operationalizations when operationalizing awareness, both in terms of enhanced internal and external validity.

The findings in the current work (Zabel & Olson, 2014) are consistent with a MODE Model perspective. Specifically, in the absence of an applicable motive (i.e., Concern), automatic attitudes directly predict social judgments. In comparison, when an applicable motive is high (i.e., Concern), the influence of automatic attitudes in social judgments can be corrected. This tendency was exemplified by high Concern individuals in the Self-Affirmation First/Introspection Second condition, for whom pro-White automatically-activated racial attitudes led to increasingly *pro-Black* trait ratings. Though findings are also generally consistent with the Moderated Consistency Model (Hofmann

et al., 2005) in that implicit-explicit divergence occurs for those with a high adjustment motive (i.e., high Concern motivation), the MODE Model provides a more comprehensive account for this and related research. Specifically, relative to the Moderated Consistency Model, the MODE Model uniquely posits that implicit-explicit consistency depends on the nature of an applicable motive. Only the MODE Model is compatible with both the current work and previous research in which motivation leads to *greater* implicit-explicit correspondence (Phillips & Olson, 2014). The current research indicates that individuals have accurate access to their automatic racial attitudes under some circumstances, but I contend that *how* this awareness prompts behavior and social judgments is dependent on the nature of the activated motive.

The current findings demonstrate the remarkable flexibility of self-affirmation in facilitating approach to otherwise self-threatening information (Harris & Napper, 2005). Indeed, self-affirmation within a realm unrelated to the nature of the self-threat was sufficient to facilitate more accurate introspection among participants who were self-affirmed and subsequently introspected. I interpret these findings as self-affirmation “opening the gate” to accurate introspection and reducing defensiveness to otherwise potentially threatening self-knowledge among highly egalitarian-oriented Whites. Indeed, converging evidence indicating the efficacy of self-affirmation in enabling attention to otherwise self-threatening information has been demonstrated using a diverse experimental procedures and methods, such as dot probe tasks (Klein & Harris, 2009), willingness to read self-threatening information (Sherman et al., 2000), and more indirectly as in the current work. These findings indicate that self-affirmations are not

only sufficient to facilitating acknowledgement of otherwise self-threatening information within health domains (Klein & Harris, 2009; Reed & Aspinwall, 1998; Sherman et al., 2000), but also within the realm of prejudice and intergroup relations (see also Adams et al., 2006).

Perhaps most prominently, the current work is consistent with an Aversive Racism perspective (Gaertner & Dovidio, 1986). An egalitarianism self-image and automatic prejudice defines aversive racists (Dovidio & Gaertner, 2004; Song Hing et al., 2008). Aversive racists engage in a variety of strategies to maintain an egalitarian self-image. Consistent with Aversive Racism Theory, the default tactic of failing to introspect on (and subsequently to correct for) one's racial prejudices among egalitarian-oriented individuals is one such strategy. The strategy is akin to classic Aversive Racism research in which liberal callers were especially quick to prematurely hang up in a phone call in conversation with an apparent Black caller in order to avoid the future prospect of their egalitarian self-image being threatened (Gaertner, 1973). Indeed, that accurate introspection and correction for automatic prejudice is avoided in the Control condition attests to the particularly insidious nature of Aversive Racism (Gaertner & Dovidio, 1986): the very individuals that perceive they are improving intergroup relations may be among the least likely of individuals to correct for their prejudices under most circumstances. Failing to introspect on their racial prejudices permits egalitarian-oriented individuals to maintain a state of "ignorant bliss," in which their egalitarian self-images are actively maintained at the expense of actual correction for automatic prejudice.

Self-affirmation appears to reduce the self-threat of introspecting on automatic racial prejudices for high egalitarian-oriented Whites, which appears to be essential in accurately identifying and subsequently correcting for automatic prejudices. In other words, self-affirmations prompted egalitarian-oriented individuals to “let their guard down” with regard to maintaining an egalitarian self-image. In such contexts, individuals were willing to actually bring to the forefront their otherwise self-threatening automatic racial attitudes when asked to introspect. These findings indicate that self-affirmation followed by introspection may reduce the tendency for aversive racists to avoid potentially self-threatening contexts that often in turn restrict prejudice correction. Indeed, self-affirmation and racial attitude introspection may curtail aversive racism effects by opening egalitarian-oriented Whites up to otherwise self-threatening information. Future research should continue to examine these possibilities. It is important to note that self-affirmation alone did not push individuals toward correcting for their automatic prejudices. If this had been the case, Self-Affirmation Only participants would also have corrected for their automatic prejudices in trait judgments, which they clearly did not. Thus, reducing the threat of accepting information potentially detrimental to one’s egalitarian self-image is only half the battle; interventions aimed at facilitating sufficient prejudice correction must also facilitate active introspection.

Convergent validity in support of the Aversive Racism tenets as theoretical underpinnings for the current findings would perhaps be most substantiated if prejudice correction effects were demonstrated toward the symbolic racism (Henry et al., 2002) explicit attitude measure among egalitarian-oriented Whites. Symbolic racism is akin to

Aversive Racism in that one's own prejudices are purported to be due to the result of value violations between one's group and another group, rather than direct hostility or a lack of egalitarianism. Indeed, just as Aversive Racism allows for maintenance of egalitarian ideals while simultaneously failing to correct for prejudices, so too does symbolic racism.

I conducted supplemental analyses with Symbolic Racism entered as the dependent variable to test for the expected automatic prejudice correction tendency among High Concern individuals with pro-White automatically-activated attitudes in the Self-Affirmation First/Introspection Second condition. A hierarchical regression uncovered the expected Predictor (Introspection First/Self-Affirmation Second vs. Self-Affirmation First/Introspection Second) x Concern x Automatically-Activated Racial Attitudes interaction, $t(191) = 3.13, p < .01$, such that among participants with pro-White automatically-activated attitudes (+1 SD) in the Self-Affirmation First/Introspection Second condition, Concern was expectedly negatively predictive of symbolic racism, $t(41) = -2.87, p < .01$. In the Self-Affirmation First/Introspection Second condition among individuals with pro-Black automatically-activated attitudes (-1 SD), Concern was not predictive of symbolic racism, $t(41) = -.20, p = .84$. Indeed, this finding, along with the automatic prejudice correction pattern among high Concern individuals in the Self-Affirmation First/Introspection Second condition regarding the prejudice composite measure and pro-Black trait ratings is congruent with an Aversive Racism perspective.

These results indicate that intergroup interaction should encourage not only egalitarianism (Rabinowitz, Wittig, Von Braun, Franke, & Zander-Music, 2005), but also

both opportunities for self-affirmations followed by honest and open personal reflection about one's own racial biases in order to more readily facilitate automatic prejudice correction. The complexities of interracial interactions, including the preferred conversational strategies (Johnson, Olson, & Fazio, 2009), expectations (Shelton, Richeson, & Salvatore, 2005), motives (Bergsieker, Shelton, & Richeson, 2010), and other psychological "baggage" that partners bring to such interactions may further complicate the degree to which self-affirmations and personal introspection of racial biases can be readily incorporated. However, given the importance of prejudice correction for intergroup interactions and lower status group member outcomes specifically, examining whether self-affirmation and racial attitude introspection elements can be successfully incorporated within interracial interactions, as well as prejudice reduction techniques more generally (Gaertner, Dovidio, Anastasio, Bachman, & Rust, 1993; Pettigrew, 1998) is warranted in future research.

In sum, the findings reported here indicate that egalitarian-oriented Whites fail to correct for their prejudices not because they lack introspective access to their racial biases, but rather because they evade introspection in order to avoid a potential threat to their egalitarian self-image. However, self-affirmation appears to reduce the self-threat of racial attitude introspection among egalitarian-oriented Whites, prompting individuals who subsequently introspected to correct for their automatic prejudices in the hypothesized manner. These results are congruent with ample research implicating the importance of self-affirmations in promoting attention the otherwise self-threatening information (Harris & Napper, 2005; Klein & Harris, 2009).

Furthermore, findings are consistent with MODE Model (Fazio, 1990) and Aversive Racism (Gaertner & Dovidio, 1986) theoretical perspectives. These findings have implications for Aversive Racism Theory. Specifically, egalitarian-oriented individuals have been demonstrated to express their prejudices when social norms are vague, as well as to engage in strategies to avoid future situations that may threaten their egalitarian self-image in previous research (Gaertner, 1973; see Pearson et al., 2009, for a review). These strategies often result in a failure to correct for prejudices, leading to negative consequences for lower status group members. The findings here indicate that self-affirmation followed by racial attitude introspection may be the specific ingredients that offer a solution to the fundamental problem of aversive racism and its negative consequences. Future research should examine whether self-affirmation and racial attitude introspection lead aversive racists to correct for prejudice even in contexts when social norms are ambiguous, as well as reduce the employment of avoidance strategies used in the service of protecting one's egalitarian self-image. Indeed, the results here are promising, and suggest that the insidious nature of and consequences stemming from aversive racism may begin to be squelched in contexts that promote self-affirmation and racial attitude introspection.

References

- Adams, G., Tormala, T. T., & O'Brien, L. T. (2006). The effect of self-affirmation on perception of racism. *Journal of Experimental Social Psychology*, 42, 616-626.
- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Armitage, C. J., Harris, P. R., Hepton, G., & Napper, L. (2008). Self-affirmation increases acceptance of health-risk information among UK adult smokers with low socioeconomic status. *Psychology of Addictive Behaviors*, 22, 88-95.
- Aronson, E. (1999). The power of self-persuasion. *American Psychologist*, 54, 875-884.
- Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 1-34). New York, NY: Academic Press.
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74, 183-200.
- Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of Personality and Social Psychology*, 99, 248-264.
- Biernat, M., & Vescio, T. K. (1993). Categorization and stereotyping: Effects of group context on memory and social judgment. *Journal of Experimental Social Psychology*, 29, 166-202.
- Blumenthal, A. (1975). A reappraisal of Wilhelm Wundt. *American Psychologist*, 30, 1081-1088.

- Bodenhausen, G. V., Macrae, G. N., & Sherman, J. W. (1999). On the dialects of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp 271-290). New York, NY: Guilford Press.
- Bodenhausen, G. V., Todd, A. R., & Richeson, J. A. (2009). Controlling prejudice and stereotyping: Antecedents, mechanisms, and contexts. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping and discrimination* (pp. 111-136). New York, NY: Psychology Press.
- Brehm, J. W. (1966) *A theory of psychological reactance*. New York: Academic Press
- Brewer, M. B. (1988). A dual-process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum.
- Carver, C. C., & Scheier, M. F. (1981). Self-consciousness and reactance. *Journal of Research in Personality*, 15, 16-29.
- Chien, Y., Wegener, D. T., Petty, R. E., & Hsiao, C. (2014). The flexible correction model: Bias correction guided by naïve theories of bias. *Social and Personality Psychology Compass*, 8, 275-286.
- Clark, J. K., Wegener, D. T., Brinol, P., & Petty, R. E. (2009). Discovering the shoe that fits: The self-validating role of stereotypes. *Psychological Science*, 20, 846-852.
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, 26, 1151-1164.

- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83, 1314-1329.
- Csikszentmihalyi, M., & Figurski, T. J. (1982). Self-awareness and aversive experience in everyday life. *Journal of Personality*, 50, 15-28.
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, 36, 1-52.
- Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11, 315-319.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice in interracial interaction. *Journal of Personality and Social Psychology*, 82, 62-68.
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, 23, 316-326.
- Duval, S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. New York, NY: Academic Press.
- Fazio, R.H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

- Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology, 18*, 339-357.
- Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology, 33*, 451-470.
- Fazio, R. H., & Hilden, L. E. (2001). Emotional reactions to a seemingly prejudiced response: The role of automatically activated racial attitudes and motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 27*, 538-549.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski & Y. Trope (Eds.), *Dual process theories of the social mind*, (pp. 155-171). New York, NY: Guilford.
- Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*, 31-44.

- Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522-527.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Fiske, S.T., Lin, M., & Neuberg, S. (1999). The continuum model: Ten years later. In Y. Trope & S. Chaiken (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York: Guilford Press.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1-74). New York: Academic Press.
- Gaertner, S. L. (1973). Helping behavior and racial discrimination among liberals and conservatives. *Journal of Personality and Social Psychology, 25*, 335-341.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-90). New York: Academic Press.
- Gaertner S. L., Dovidio J. F., Anastasio, P. A., Bachman B. A., & Rust M. C. (1993). The common in-group identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology, 4*, 1-26.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.

- Gawronski, B., & LeBel, E. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44*, 1355-1361.
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*, 617-638.
- Greenberg, J., & Pyszczynski, T. (1985). The effect of an overheard ethnic slur on evaluations of the target: How to spread a social disease. *Journal of Experimental Social Psychology, 21*, 61-72.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (1995). Understanding and using the Implicit Association Test: An improved algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.
- Gschwendner, T., Hofmann, W., & Schmitt, M. (2006). Synergistic moderator effects of situation and person factors of awareness and adjustment on the consistency of implicit and explicit attitudes. *Journal of Individual Differences, 27*, 47-56.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review, 103*, 336-355.
- Harris, P. R., Mayle, K., Mabbott, L., & Napper, N. (2007). Self-affirmation reduces smokers' defensiveness to graphic on-pack cigarette labels. *Health Psychology, 26*, 437-446.

- Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of threatening health-risk information. *Personality and Social Psychology Bulletin*, *31*, 1250-1263.
- Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, *23*, 253-283.
- Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*, 319-340.
- Hixon, J. G., & Swann, W. B., Jr. (1993). When does introspection bear fruit? *Journal of Personality and Social Psychology*, *64*, 35-43.
- Hodges, S. D., & Wilson, T. D. (1993). Effects of analyzing reasons on attitude change: The moderating role of attitude accessibility. *Social Cognition*, *11*, 353-366.
- Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychological Bulletin*, *28*, 460-471.
- Hofmann, W., Gschwendner, T., Nosek, B. A., Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology*, *16*, 335-390.
- Hunt, M. (2007). *The Story of Psychology*. New York, NY: Random House.
- James, W. (1890). *The Principles of Psychology* (Vol. 1). New York, NY: Henry Holt & Company.

- Johnson, C. S., Olson, M. A., & Fazio, R. H. (2009). Getting acquainted in interracial interactions: Avoiding intimacy but approaching race. *Personality and Social Psychology Bulletin*, 35, 557-571
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, 55, 893-905.
- Kirschbaum, C., Pirke, K., & Hellhammer, D. H. (1993). The trier social stress test: A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28, 76-81.
- Klein, W. M. P., & Harris, P. R. (2009). Attentional bias toward threatening components of a persuasive message. *Psychological Science*, 20, 1463-1467.
- Koole, S. L., Smeets, K., van Knippenberg, A., & Dijksterhuis, A. The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, 77, 111-125.
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, 103, 284-308.
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51, 93-120.
- McCarthy, R. J., & Skowronski, J. J. (2014). Controlled processing and automatic processing in the formation of spontaneous trait inferences. In J. W. Sherman, B.

- Gawronski, & Y. Trope (Eds), *Dual-process theories of the social mind* (pp. 221-234). New York, NY: Guilford Press.
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association task, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435-442.
- McNulty, J. K., Olson, M. A., Meltzer, A. L., & Shaffer, M. J. (2013). Though they may be unaware, newlyweds implicitly know whether their marriage will be satisfying. *Science*, 342, 1119–1120.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81, 33-43.
- Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, 65, 469-485.
- Monteith, M. J., & Mark, A. Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. *European Review of Social Psychology*, 16, 113-154.
- Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, 19, 395-417.
- Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, 77, 167-184.

- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Olson, M. A., & Fazio, R. H. (2004). Trait inferences as a function of automatically activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology*, 26, 1-11.
- Payne, B.K., Cheng, S.M., Govorun, O., & Stewart, B.D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277-293.
- Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, 3, 314-338.
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65-85.
- Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, 25, 57-75.
- Petty, R. R., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 123-205). New York, NY: Guilford Press.
- Phillips, J. E., & Olson, M. A. (2014). When implicitly and explicitly measured racial attitudes align: The roles of social desirability and thoughtful responding. *Basic and Applied Social Psychology*, 36, 125-132.

- Plant, E. A. (2004). Responses to interracial interactions over time. *Personality and Social Psychology Bulletin*, 30, 1458-1471.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832.
- Plant, E. A., Devine, P. G., & Peruche, M. B. (2010). Routes to positive interracial interactions: Approaching egalitarianism or avoiding prejudice. *Personality and Social Psychology Bulletin*, 36, 1135-1147.
- Rabinowitz, J. L., Wittig, M. A., von Braun, M., Franke, R., & Zander-Music, L. (2005). Understanding the relationship between egalitarianism and affective bias: Avenues to reducing prejudice among adolescents. *Journal of Social Issues*, 51, 525-545.
- Reed, M. B., & Aspinwall, L. G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion*, 22, 99-132.
- Sanbonmatsu, D. M., & Fazio, R. H. (1990). The role of accessibility in memory-based decision-making. *Journal of Personality and Social Psychology*, 59, 614-622.
- Schuetz, R. A., & Fazio, R. H. (1995). Attitude accessibility and motivation as determinants of biased processing: A test of the MODE model. *Personality and Social Psychology Bulletin*, 21, 704-710.
- Sedikides, C., Horton, R. S., & Gregg, A. P. (2007). The why's the limit: Curtailing self-enhancement with explanatory introspection. *Journal of Personality*, 75, 783-824.

- Shelton, J. N., Richeson, J. A., & Salvatore, J. (2005). Expecting to be the target of prejudice: Implications for interethnic interactions. *Personality and Social Psychology Bulletin*, 31, 1189-1202.
- Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychological Science*, 16, 397-402.
- Sherman, D. K., Bunyan, D. P., Creswell, J. D., & Jaremka, L. M. (2009). Psychological vulnerability of stress: The effects of self-affirmation on sympathetic nervous system responses to naturalistic stressors. *Health Psychology*, 28, 554-562.
- Sherman, J. W., Gawronski, B., & Trope, Y. (2014). *Dual-process theories of the social mind*. New York, NY: Guilford Press.
- Sherman, D. K., and Hartson, K. A. (2011). Reconciling self-defense with self criticism: Self-affirmation theory. In M. D. Alicke & C. Sedikides (Eds). *Handbook of self-enhancement and self-protection* (pp. 128–151). New York, NY: Guilford Press.
- Sherman, D. K., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin*, 26, 1046-1058.
- Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, 77, 885.
- Son Hing, L. S., Chung-Yan, G., Hamilton, L., & Zanna, M. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology*, 94, 971–987.

- Steele, C. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 181-227). San Diego, CA: Academic Press.
- Steele, C. M., & Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology*, 45, 5-19.
- Steele, C., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology*, 64, 885-896.
- Strack, F., & Deutsch, R. (2014). The reflective-impulsive model. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 92-104). New York, NY: Guilford.
- Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54, 768-777.
- Tesser, A., & Cornell, D. P. (1991). On the confluence of self processes. *Journal of Experimental Social Psychology*, 27, 501-526.
- Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically activated racial attitudes and anticipated comfort interacting with African Americans. *Personality and Social Psychology Bulletin*, 29, 170-182.

- Towles-Schwen, T., & Fazio, R. H. (2001). On the origins of racial attitudes: Correlates of childhood experiences. *Personality and Social Psychology Bulletin*, 27, 162-175.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). San Diego, CA: Academic Press.
- Wegener, D.T., Petty, R.E., & Dunn, M. (1998). The metacognition of bias correction: Naive theories of bias and the flexible correction model. In V.Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Meta-cognition: Cognitive and social dimensions* (pp. 202–227). London: Sage.
- Wells, G. L., & Petty, R. E. (1980). The effects of over head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology*, 1, 219-230.
- Wilson, T. D. (2009). Know thyself. *Perspectives on Psychological Science*, 4, 384-389.
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117-142.
- Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology*, 55, 493-518.

- Wilson, T. D., Dunn, D., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude-behavior consistency. *Journal of Personality and Social Psychology*, 47, 5-16.
- Wilson, T. D., Dunn, D., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). Orlando, FL: Academic Press.
- Wilson, T. D., Kraft, D., & Dunn, D. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology*, 25, 379-400.
- Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., & Klaaren, S. J., et al. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin*, 19, 331-339.
- Zabel, K. L., & Olson, M. A. (2014). Unpublished Manuscript. The operation of impression-relevant motives depends on construal level of social targets.

Appendices

Appendix A: Tables

Table 1. Descriptive Statistics and Correlations Between Variables For All Participants (Averaging Across Condition)

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Concern Motivation	.00	.95	---											
2. Restraint Motivation	-.03	.98	.05	---										
3. Automatic Racial Attitudes ^a	.000003	.00012	-.10	-.11	---									
4. Pro-Black Trait Ratings	.25	.31	.07	.00	-.03	---								
5. Pro-Black Individual Ratings	.09	.35	.11	.07	-.04	.84**	---							
6. Pro-Black Group Ratings	.45	.37	.07	-.03	-.07	.85**	.44**	---						
7. Prejudice Composite	.000	.83	-.36**	-.02	.09	-.31**	-.23**	-.39**	---					
8. Subtle Prejudice	2.39	.58	-.32**	-.09	.13*	-.29**	-.26**	-.31**	.89**	---				
9. Blatant Racism	2.03	.54	-.25**	-.01	.03	-.31**	-.19**	-.40**	.82**	.65**	---			
10. Symbolic Racism	2.34	.48	-.35**	-.02	.06	-.21**	-.12*	-.30**	.85**	.70**	.62**	---		
11. Anti-Black Attitudes	3.81	.74	-.28**	.05	.08	-.23*	-.19**	-.26**	.76**	.60**	.43**	.48**	---	
12. Pro-Black Attitudes	4.18	.69	.35**	.03	-.03	.07	.04	.15**	-.41**	-.39**	-.36**	-.52**	-.08	---
13. Gender ^b	---	---	-.33**	-.13*	.05	.01	-.00	-.02	.16**	.17**	.09	.19**	.05	-.04

** Correlations are significant at $p < .05$; * Correlations are significant at $p < .10$

^a Higher scores indicate more automatically prejudiced attitude; ^b = Female = 0; Male = 1

Table 2. Descriptive Statistics and Correlations Between Variables For Control Condition Participants

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Concern Motivation	-.25	1.00	---											
2. Restraint Motivation	.18	1.19	.21	---										
3. Automatic Racial Attitudes ^a	.000004	.00013	-.10	-.27*	---									
4. Pro-Black Trait Ratings	.26	.39	.26*	-.17	-.06	---								
5. Pro-Black Individual Ratings	.07	.46	.24	-.16	.01	.90**	---							
6. Pro-Black Group Ratings	.45	.42	.22	-.13	-.12	.88**	.59**	---						
7. Prejudice Composite	-.02	.85	-.46**	.04	.02	-.66**	-.65**	-.51**	---					
8. Subtle Prejudice	2.36	.59	-.45**	.04	.06	-.63**	-.69**	-.42**	.91**	---				
9. Blatant Racism	2.05	.53	-.36**	.06	-.14	-.55**	-.46**	-.52**	.79**	.64**	---			
10. Symbolic Racism	2.34	.51	-.38**	-.04	.03	-.49**	-.50**	-.36**	.85**	.71**	.67**	---		
11. Anti-Black Attitudes	3.78	.80	-.34**	.06	.10	-.50**	-.50**	-.39**	.74**	.65**	.33**	.44**	---	
12. Pro-Black Attitudes	4.22	.69	.46**	-.10	-.10	.45**	.46**	.33**	-.58**	-.54**	-.53**	-.65**	-.21	---
13. Gender ^b	---	---	-.43**	-.18	.03	.05	.08	.02	.32**	.32**	.28*	.35**	.11	-.13

** Correlations are significant at $p < .05$; * Correlations are significant at $p < .10$

^a Higher scores indicate more automatically prejudiced attitude; ^b = Female = 0; Male = 1

Table 3. Descriptive Statistics and Correlations Between Variables For Introspection Only Condition Participants

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Concern Motivation	-.06	.81	---											
2. Restraint Motivation	-.08	.89	.34**	---										
3. Automatic Racial Attitudes ^a	-.00004	.00013	.01	-.13	---									
4. Pro-Black Trait Ratings	.13	.25	-.20	-.05	-.09	---								
5. Pro-Black Individual Ratings	.06	.28	-.06	.10	-.12	.82**	---							
6. Pro-Black Group Ratings	.32	.35	-.04	-.04	-.09	.81**	.21	---						
7. Prejudice Composite	-.09	.78	-.47**	-.23	-.11	-.18	-.09	-.63**	---					
8. Subtle Prejudice	2.38	.55	-.47**	-.37**	-.06	-.17	-.16	-.54**	.91**	---				
9. Blatant Racism	1.91	.47	-.33**	-.32**	.05	-.26*	-.08	-.67**	.78**	.67**	---			
10. Symbolic Racism	2.36	.47	-.42**	.15	-.12	-.01	.05	-.35	.88**	.80**	.56**	---		
11. Anti-Black Attitudes	3.69	.71	-.35**	.03	-.21	-.19	-.10	-.45**	.76**	.55**	.40**	.55**	---	
12. Pro-Black Attitudes	3.98	.78	.45**	.35**	.06	-.31**	-.20	.05	-.43**	-.46**	-.26*	-.52**	-.17	---
13. Gender ^b	---	---	-.23**	-.06	.05	.06	.10	-.26	.34**	.32**	.29*	.40**	.13	-.16

** Correlations are significant at $p < .05$; * Correlations are significant at $p < .10$

^a Higher scores indicate more automatically prejudiced attitude; ^b = Female = 0; Male = 1

Table 4. Descriptive Statistics and Correlations Between Variables For Self-Affirmation Only Condition Participants

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Concern Motivation	.22	1.03	---											
2. Restraint Motivation	-.12	.97	-.16	---										
3. Automatic Racial Attitudes ^a	.000006	.00012	-.07	-.05	---									
4. Pro-Black Trait Ratings	.26	.28	-.03	.01	.07	---								
5. Pro-Black Individual Ratings	.08	.31	.11	.10	.05	.78**	---							
6. Pro-Black Group Ratings	.45	.38	-.13	-.07	.06	.86**	.36**	---						
7. Prejudice Composite	.05	.93	-.46**	.08	-.02	-.28*	-.19	-.27*	---					
8. Subtle Prejudice	2.47	.67	-.29*	-.04	-.10	-.31**	-.21	-.29*	.86**	---				
9. Blatant Racism	2.10	.60	-.36**	.04	.10	-.31**	-.20	-.30**	.89**	.71**	---			
10. Symbolic Racism	2.31	.52	-.49**	.16	.04	-.07	.02	-.11	.84**	.63**	.66**	---		
11. Anti-Black Attitudes	3.80	.74	-.42**	.12	-.13	-.27*	-.26*	-.20	.80**	.55**	.62**	.55**	---	
12. Pro-Black Attitudes	4.25	.69	.53**	-.24	-.19	-.18	-.11	-.19	-.41**	-.38**	-.37**	-.48**	-.13	---
13. Gender ^b	---	---	-.05	-.05	.06	.14	.16	.08	-.15	-.09	.18	.00	-.25*	-.01

Correlations are significant at $p < .05$; * Correlations are significant at $p < .10$

^a Higher scores indicate more automatically prejudiced attitude; ^b = Female = 0; Male = 1

Table 5. Descriptive Statistics and Correlations Between Variables For Introspection First/Self-Affirmation Second Condition Participants

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Concern Motivation	.20	.77	---											
2. Restraint Motivation	-.14	.93	-.08	---										
3. Automatic Racial Attitudes ^a	.000013	.00010	-.10	-.16	---									
4. Pro-Black Trait Ratings	.32	.32	-.05	.32*	-.12	---								
5. Pro-Black Individual Ratings	.12	.42	-.06	.44**	-.14	.89**	---							
6. Pro-Black Group Ratings	.51	.33	-.02	.05	-.07	.81**	.44**	---						
7. Prejudice Composite	.02	.69	-.07	-.07	.31*	.01	.20	-.24	---					
8. Subtle Prejudice	2.37	.48	-.16	-.13	.36**	.09	.19	-.07	.89**	---				
9. Blatant Racism	2.06	.50	-.09	-.02	.28*	-.14	.04	-.33**	.78**	.61**	---			
10. Symbolic Racism	2.29	.44	-.00	-.10	.14	-.06	.14	-.30*	.79**	.64**	.64**	---		
11. Anti-Black Attitudes	3.93	.73	.02	.04	.17	.13	.22	-.02	.60**	.50**	.14	.15	---	
12. Pro-Black Attitudes	4.24	.68	.05	.15	-.03	-.03	-.17	.15	-.26	-.18	-.41**	-.59**	.36**	---
13. Gender ^b	---	---	-.24	-.35**	.15	-.13	-.16	-.04	.38**	.50**	.25	.14	.27*	.04

** Correlations are significant at $p < .05$; * Correlations are significant at $p < .10$

^a Higher scores indicate more automatically prejudiced attitude; ^b = Female = 0; Male = 1

Table 6. Descriptive Statistics and Correlations Between Variables For Self-Affirmation First/Introspection Second Condition Participants

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. Concern Motivation	-.09	1.04	---											
2. Restraint Motivation	.02	.88	.07	---										
3. Automatic Racial Attitudes ^a	.000034	.00013	-.22	.03	---									
4. Pro-Black Trait Ratings	.27	.26	.18	.04	-.16	---								
5. Pro-Black Individual Ratings	.10	.27	.18	.01	-.08	.81**	---							
6. Pro-Black Group Ratings	.45	.33	.13	.04	-.18	.88**	.43**	---						
7. Prejudice Composite	.04	.87	-.31**	.03	.28*	-.37**	-.24	-.37**	---					
8. Subtle Prejudice	2.36	.58	-.30**	.00	.30**	-.27*	-.25	-.21	.90**	---				
9. Blatant Racism	2.04	.56	-.16	.11	.05	-.35**	-.20	-.37**	.80**	.59**	---			
10. Symbolic Racism	2.37	.50	-.34**	-.02	.24	-.32**	-.13	-.38**	.88**	.74**	.60**	---		
11. Anti-Black Attitudes	3.88	.73	-.29*	.00	.38**	-.32**	-.24	-.30**	.86**	.77**	.53**	.66**	---	
12. Pro-Black Attitudes	4.23	.58	.19	.00	-.05	.17	.04	.23	-.42**	-.39**	-.37**	-.41**	-.27*	---
13. Gender ^b	---	---	-.63**	-.08	-.04	-.11	-.21	-.00	.03	.01	-.04	.07	.06	.09

** Correlations are significant at $p < .05$; * Correlations are significant at $p < .10$

^a Higher scores indicate more automatically prejudiced attitude; ^b = Female = 0; Male = 1

Appendix B: Figures

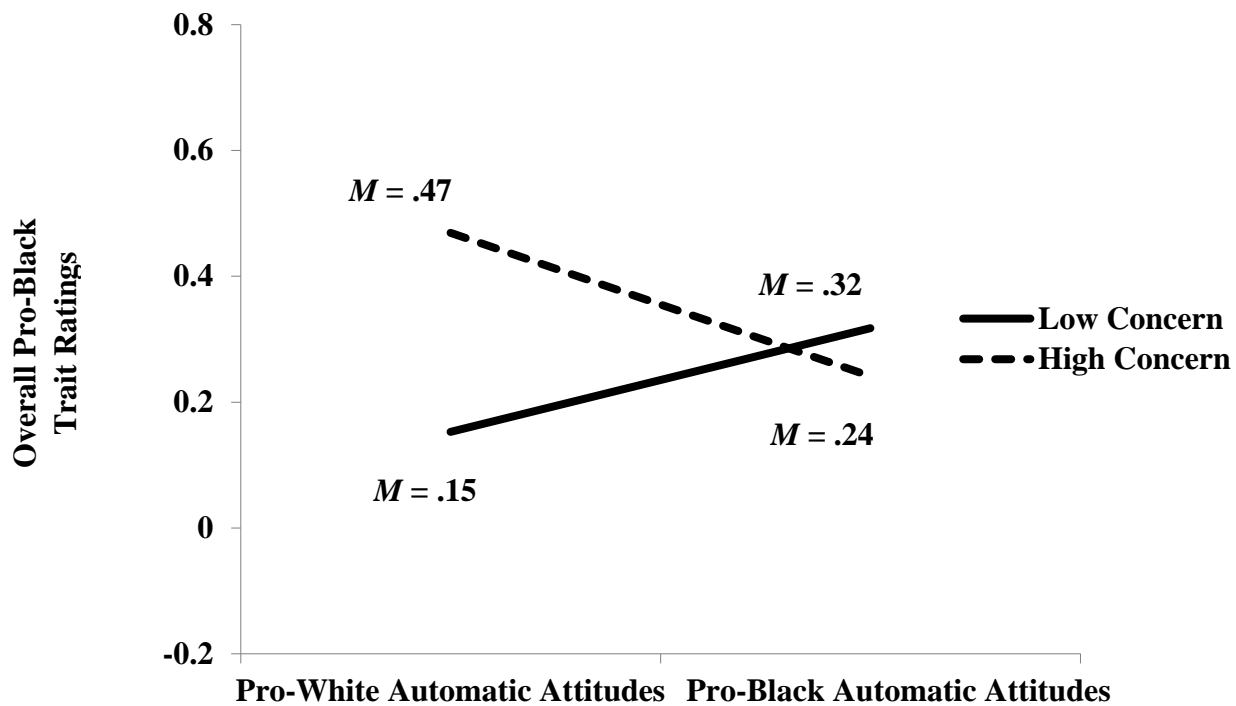


Figure 1. Concern Motivation x Automatically-Activated Racial Attitudes interaction in predicting overall pro-Black trait ratings in the Self-Affirmation First/Introspection Second condition

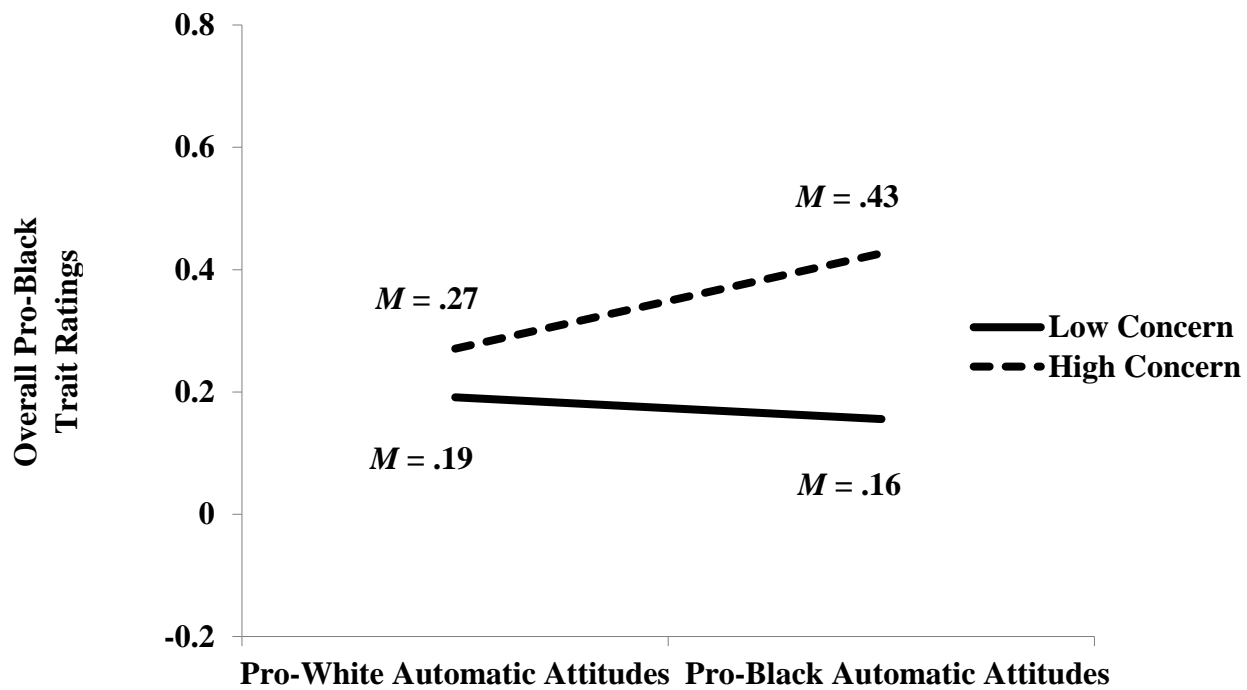


Figure 2. Concern Motivation x Automatically-Activated Racial Attitudes interaction in predicting overall pro-Black trait ratings in the Control condition

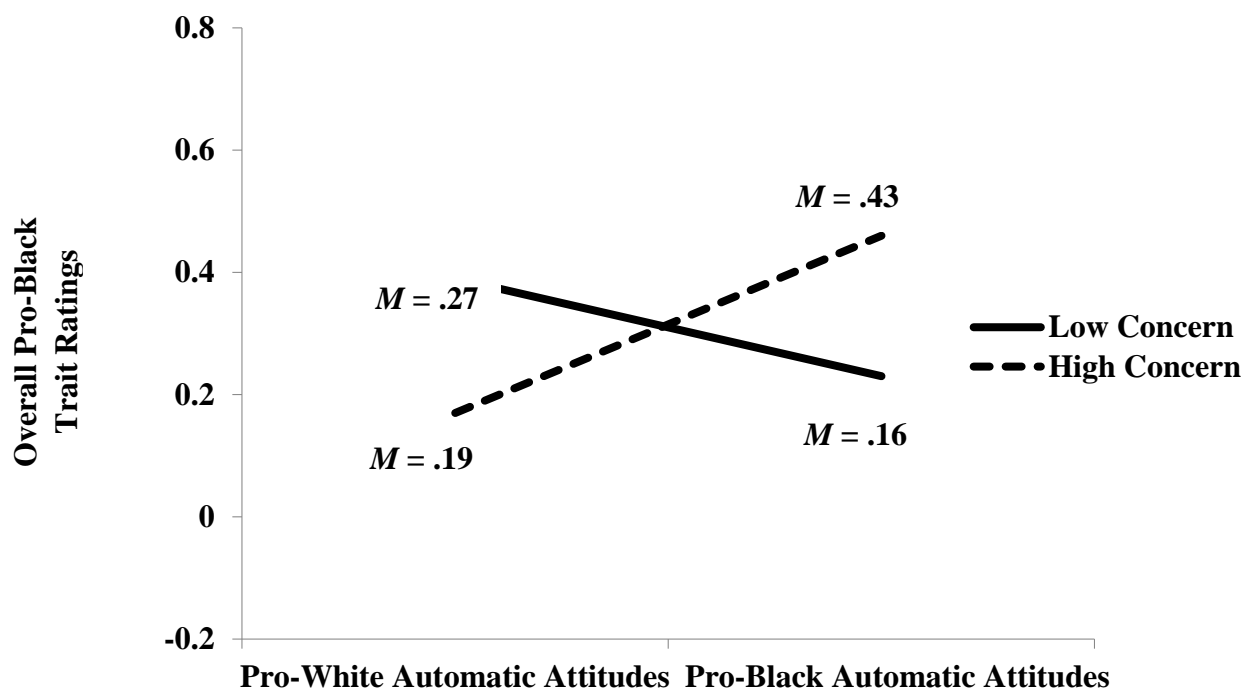


Figure 3. Concern Motivation x Automatically-Activated Racial Attitudes interaction in predicting overall pro-Black trait ratings in the Introspection First/Self-Affirmation Second condition

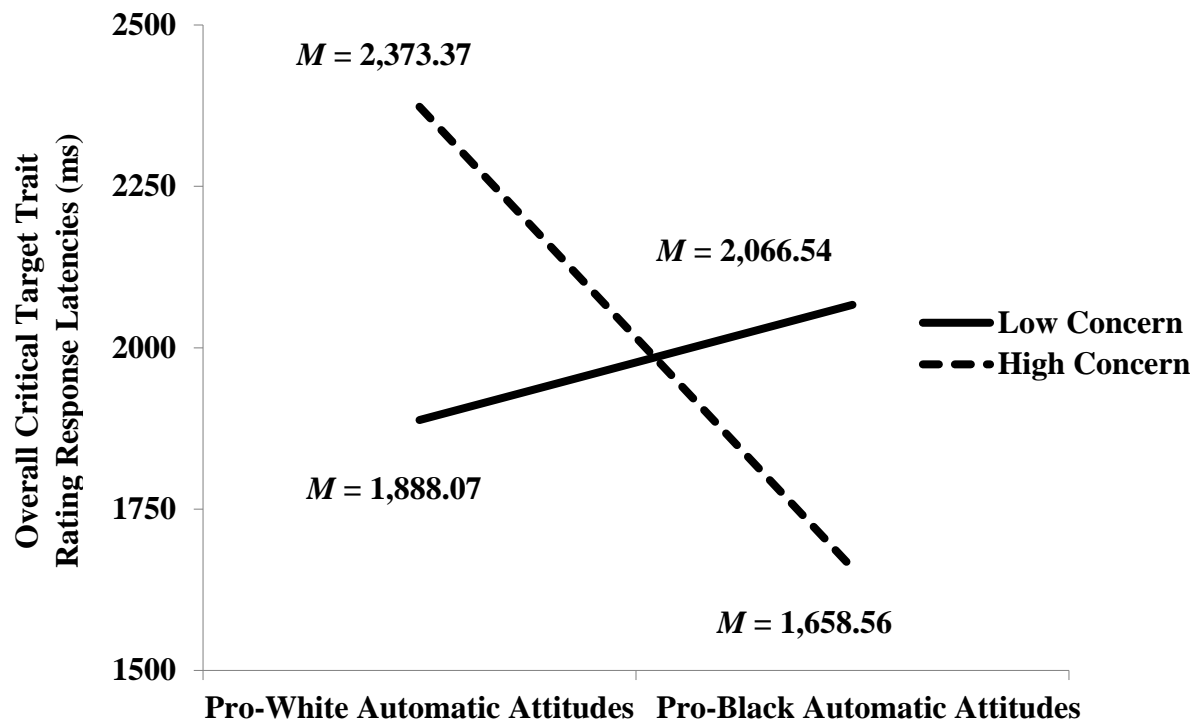


Figure 4. Concern Motivation x Automatically-Activated Racial Attitudes interaction in predicting mean overall critical target trait rating response latencies in the Self-Affirmation First/Introspection Second condition

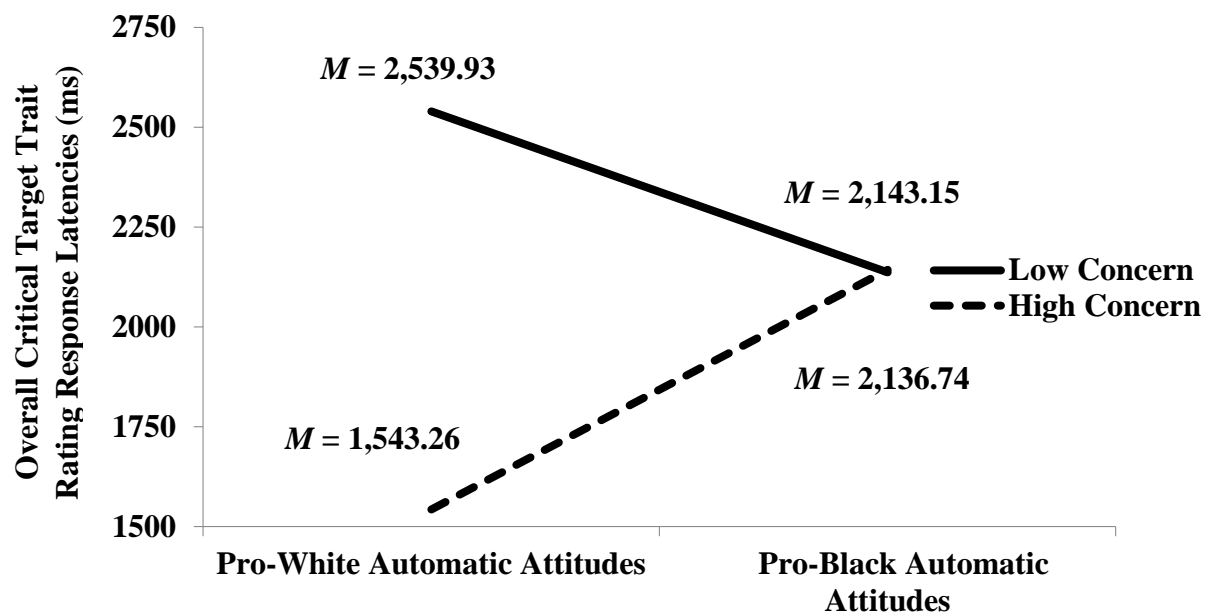


Figure 5. Concern Motivation x Automatically-Activated Racial Attitudes interaction in predicting overall mean critical target trait rating response latencies in the Control condition

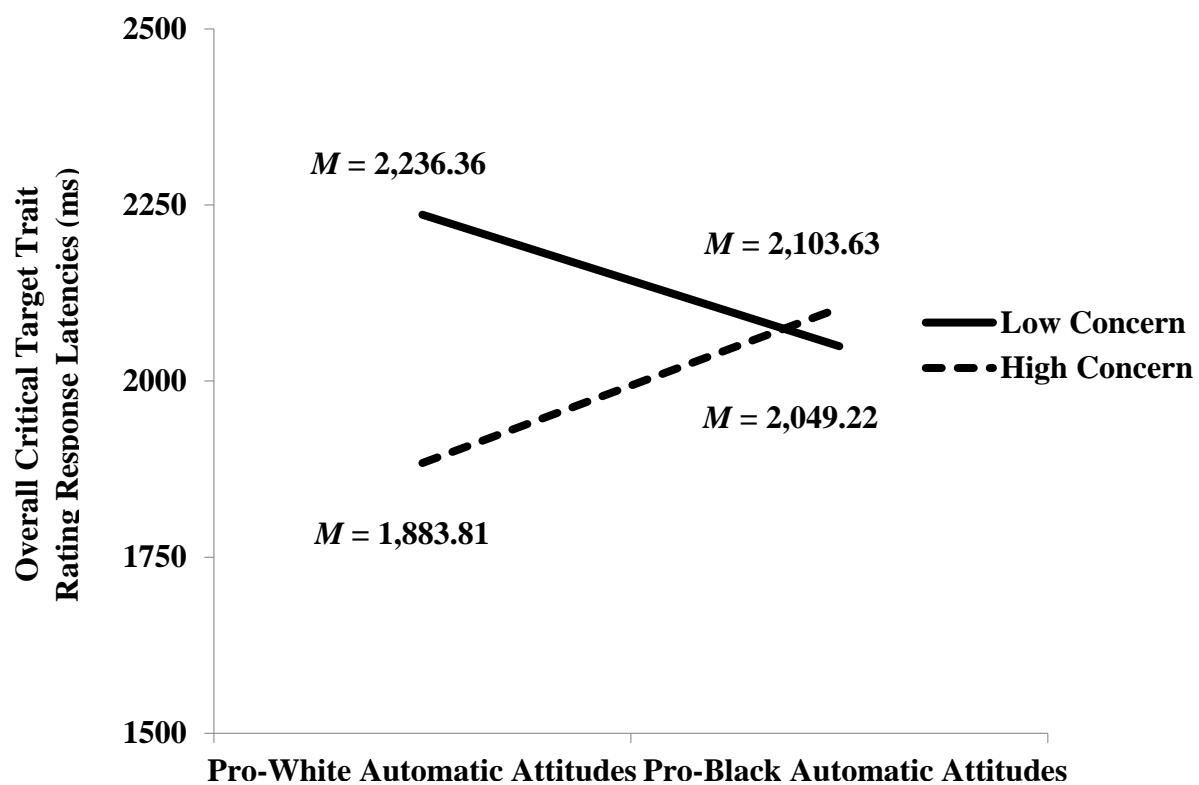


Figure 6. Concern Motivation x Automatically-Activated Racial Attitudes interaction in predicting overall mean critical target trait rating response latencies in the Self-Affirmation Only condition

VITA

Kevin Zabel was born in Three Oaks, MI, to the parents of Chuck and Tammy Zabel, and to the grandparents of Chuck Zabel Sr. and Ruth Ann Zabel, as well as Donald and Donna Siegmund. He is a brother to his identical twin brother, Keith, and younger sister, Nicole. He attended Three Oaks Elementary School and graduated from River Valley High School in Three Oaks, MI, in 2005. After graduation, he attended Albion College in Albion, MI, where he double-majored in Psychology and Economics and Management. Under the mentorship of Andrew Christopher, he developed a passion for psychological research. After obtaining his Bachelor of Arts degree in Psychology and Economics and Management, he worked for one year as the Academic Program Coordinator and academic class assistant for Scott Chesebro at the Chicago Center for Urban Life and Culture, an urban studies program for undergraduate college students in the Hyde Park neighborhood of Chicago. Following his apprenticeship, he accepted a graduate research fellowship at the University of Tennessee in Knoxville, TN. For five years, he worked under the mentorship of Michael Olson, conducting research on social cognition, automatic attitudes, interracial interactions, and the dynamics of racial prejudice. At the University of Tennessee, he developed a deepened passion not only for conducting psychological research, but also for teaching and working with students in the classroom and research lab. He graduated with his Ph.D. in Experimental Psychology (Social concentration) in December 2015. Kevin recently started a tenure-track assistant professor position in the Department of Psychology at Western New England University in Springfield, MA in August 2015. Kevin will teach classes including Research

Methods, Statistics for the Behavioral Sciences, Introductory Psychology, and Cognitive Psychology. Already, Kevin has developed a fondness for his fellow faculty members and students, and he is excited to begin the next chapter of his professional life.